

ON HAVING MEANING IN MIND

Jesper Kallerstrup

**A Thesis Submitted for the Degree of PhD
at the
University of St Andrews**



2001

**Full metadata for this item is available in
St Andrews Research Repository
at:**

<http://research-repository.st-andrews.ac.uk/>

Please use this identifier to cite or link to this item:

<http://hdl.handle.net/10023/12911>

This item is protected by original copyright

On Having Meaning in Mind

Jesper Kallestrup

May 2001

Ph.D. Thesis

Department of Logic and Metaphysics

University of St. Andrews



ProQuest Number: 10170793

All rights reserved

INFORMATION TO ALL USERS

The quality of this reproduction is dependent upon the quality of the copy submitted.

In the unlikely event that the author did not send a complete manuscript and there are missing pages, these will be noted. Also, if material had to be removed, a note will indicate the deletion.



ProQuest 10170793

Published by ProQuest LLC (2017). Copyright of the Dissertation is held by the Author.

All rights reserved.

This work is protected against unauthorized copying under Title 17, United States Code
Microform Edition © ProQuest LLC.

ProQuest LLC.
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106 – 1346

π
D 860

Declarations

(i) I, Jesper Kallestrup, hereby certify that this thesis, which is approximately 86,000 words in length, has been written by me, that it is the record of work carried out by me and that it has not been submitted in any previous application for a higher degree.

date 26/5/01..... signature of candidate

(ii) I was admitted as a research student in September 1997 and as a candidate for the degree of PhD in September 1997; the higher study for which this is a record was carried out in the University of St. Andrews between 1997-2001.

date 26/5/01..... signature of candidate

(iii) I hereby certify that the candidate has fulfilled the conditions of the Resolutions and Regulations appropriate for the degree of PhD in the University of St. Andrews and that the candidate is qualified to submit this thesis in application for that degree.

date 17.5.2001..... signature of supervisor

(iv) In submitting this thesis to the University of St. Andrews I understand that I am giving permission for it to be made available for use in accordance with the regulations of the University Library for the time being in force, subject to any copyright vested in the work not being affected thereby. I also understand that the title and abstract will be published, and that any work may be made and supplied to any bona fide library or research worker.

date 26/5/01..... signature of candidate

"If God had looked into our minds he would not have been able to see there whom we were speaking of."

(Wittgenstein, *Philosophical Investigations*)

Preface

There are many people who, in one way or another, have had an influence on this thesis. There is the famous Southgait Group from whose members - Lars Binderup, Patrick Greenough, Lars Gundersen, Patrice Philie, Duncan Prithcard and Sven Rosenkrantz - I have learned so much. After many hours of discussion over the last three years, I have not only received invaluable criticism of my own work, I have also gained insight from their research into other aspects of analytical philosophy. I am, moreover, grateful to the philosophers at the Department of Logic & Metaphysics, University of St. Andrews, for the many useful comments they have given me in research seminars, conversations and reviews. In particular, I would like to thank the members of my review board Fraser MacBride and Steve Read.

In the spring of 2000 I was fortunate to be able to study at the Research School of Social Sciences, Australian National University, for a period of two months. The combination of an extremely sociable environment, together with some of the best permanent as well as visiting philosophers in the world, makes its Philosophy Programme an ideal place to do philosophy. As readers of this thesis will discover, this place has played a significant role in shaping my philosophical views.

Most of all I am indebted to my supervisor Crispin Wright. If anyone has taught me how to do philosophy, it is Crispin. I have learned never to rest content, but always to strive to improve a view by making the contrasting view as strong as possible. He has been a source of constant inspiration and made it so much fun to do philosophy. Most importantly, support and encouragement have always been available when needed.

Finally, I should like to thank the Danish Research Academy for the generous stipend which made the undertaking of this PhD possible in the first place.

Abstract

There have traditionally been two views as to what makes it the case that a singular term has the propositional content that it does. According to Descriptivism, the content of a term is descriptive since it is given by a cluster of descriptive properties commonly associated with it. According to Referentialism, the content of a term is singular since it is determined by the object it picks out. It follows that empty terms can have descriptive, but not singular, content. If narrow content is what intrinsic duplicates have in common, then descriptive content is arguably narrow. Singular content, however, is wide since intrinsic duplicates who inhabit different environments express different singular contents by the same terms. On the face of it, the arguments against Descriptivism and Semantic Internalism - the view that content is narrow - seem convincing, but a worry persists, namely how to reconcile Semantic Externalism - the view that content is wide - with the kind of Privileged Access speakers enjoy with respect to the contents of their occurrent attitudes. By thorough examination of those arguments I find space for an intermediate position. What we learn is not that reference cannot go by properties, but rather which properties mediate reference. Kripke's Modal Argument proves that we need rigidified descriptive properties, Putnam's Twin Earth Argument shows that we better include causal properties, and Burge's Arthritis Argument highlights that we frequently invoke properties involving reference to other speakers. What is more, considerations about the behaviour of singular terms in intentional contexts strongly suggest that their propositional contents cannot be exhausted by their referents. By deploying so-called Two-Dimensionalism, as developed by Stalnaker, Kaplan, Evans, Davies, and others, I argue that singular content is knowable only after relevant empirical information about the actual world is in, and so is not subject to Privileged Access. Descriptive content, however, is a priori knowable since it is independent of which possible world is actual. But if that is so, then descriptive content constrained by rigidity, causality and other-dependence must also be a priori. All it takes is knowledge of how to describe various thought experiments. Although the latter kind of content is not object-dependent, as is singular content, it is wide in that it fails to be shared by duplicates who are embedded in distinct physical or social environments. So, we should expect compatibility between Semantic Externalism and Privileged Access only if Rigidified Causal Descriptivism is adopted.

Table of Contents

Introduction	1
Chapter 1. Incompatibilism	12
1.1 Semantic Externalism	12
1.2. The Twin-Earth Argument	13
1.3. Privileged Access	22
1.4. Incompatibilism	27
1.5. The MC-form	36
1.6. Weak or Strong Semantic Externalism	39
Chapter 2. The Modal Argument	47
2.1. Semantic Content and Mental Content	47
2.2. Semantic Internalism and Descriptivism	48
2.3. Semantic Externalism and Referentialism	56
2.4. The Modal Argument	64
2.5. Rigidification	76
2.6. The Rigidity Principle	81
Chapter 3. The Intentional Argument	91
3.1. Substitutivity	91
3.2. Semantics or Pragmatics	98
3.3. The Intentional Argument	104
3.4. Kripke's Paradox	108
3.5. The Strengthened Kripke's Paradox	114
3.6. Communication	122

Chapter 4. Two-Dimensionalism	131
4.1. Causal Rigidified Descriptivism	131
4.2. The Logic of 'Actually'	138
4.3. Two-Dimensionalism	150
4.4. Two Grades of Understanding	157
4.5. The Contingent A Priori	168
4.6. The Necessary A Posteriori	177
 Chapter 5. Beyond Incompatibilism	 185
5.1. Three Kinds of External Dependence	185
5.2. What is Narrow Content?	192
5.3. Some Objections to Narrow Content	201
5.4. The Empty Case Revisited	210
5.5. A Response to the MC-form	215
5.6. Beyond Incompatibility	220
 Bibliography	 226

Introduction

It used to be widely acknowledged that meanings were in the mind. What an expression meant was a question about what a competent speaker who understood that expression could be said to know just by introspective reflection. To grasp the meaning of a referring expression was to know a set of associated descriptive conditions that an object had to satisfy uniquely in order to be the referent of that expression. No expression could refer to an object unless it had certain descriptive properties knowledge of which constituted a speaker's grasp of the meaning of the expression. It was no part of the meaning of an expression which object, if any, it referred to, and so, no part of understanding such an expression that one had knowledge of its reference. The relation between an expression and its referent was thus essentially mediated by meaning: to understand a referring expression was a matter of knowing the relation between the expression and its associated descriptive meaning, and for an object to be the referent of the expression was a matter of satisfying the associated meaning. The mind associated certain descriptive conditions with an expression independently of any facts external to the mind, and then the expression had a reference if and only if there was an object which uniquely met those conditions. Meanings were safely in the mind and it was up to the objects in the world to satisfy them in order to be picked out by expressions having those meanings.

This *Descriptivist* picture of meaning and reference goes back to Frege and Russell, is often ascribed to Strawson, Searle and Dummett, and has nowadays persistently been defended by David Lewis and Frank Jackson. It came under heavy attack a couple of decades ago by arguments put forward by Kripke, Putnam, Donnellan and others. According to this new *Referentialist* picture, meanings were now no longer thought of in isolation from facts external to the mind, but as determined by the nature of the referents of expressions having those meanings. Referring expressions were directly hooked up with their referents unmediated by associated conditions. This meant that the meaning of an

expression was exhausted by its referent such that to know its meaning was to know which object it referred to. To understand an expression was to know its reference. Knowledge of the meaning of a referring expression was no longer equated with reflectively accessible knowledge of certain mental associations. To know the meaning of an expression was to know which worldly object it picked out and such De Re knowledge was taken to require the obtaining of some appropriate causal, or other epistemically relevant, external link between the speaker and the object.

The consequences were profound. If semantic contents were determined by facts external to the mind, then so were mental contents which are the objects of our propositional attitudes. Both Descriptivists and Referentialists have agreed that certain consequences about mental content followed from the theory of reference and meaning. According to Descriptivism, the propositional contents of cognitive attitudes are purely *conceptual propositions* consisting of the descriptive meanings under which the objects of those attitudes are presented in thought. The descriptive contents of such attitudes are *narrow* in the sense of being individuated by reference to what goes on only *inside* the individual who has those attitudes. To have an attitude with a narrow content is to have a mental property that is necessarily shared by *intrinsic* duplicates, or *doppelgängers*, as we shall say. For what marks the difference between doppelgängers is their *relational* properties, their different physical or social history. On the Referentialist account, however, the propositional contents of cognitive attitudes are *singular propositions* consisting of, or individuated by, the objects and properties themselves that are the referents of the embedded content clauses. The singular contents of attitudes are *wide* in the sense of being individuated by reference to what goes on also *outside* the individual who has those attitudes. To have an attitude with a wide content is thus to have a mental property that is not necessarily shared by such doppelgängers. In short, Descriptivism and Referentialism in the philosophy of language have traditionally been accompanied by *Semantic Internalism* and *Semantic Externalism* respectively in the philosophy of mind.

I think it is fair to say that the philosophical community, almost without exception, was taken in by the arguments brought forward in support of Referentialism and Semantic Externalism - indeed to such an extent that those doctrines nowadays rank as the received view. But a worry has persisted, namely that of reconciling Semantic Externalism with the kind of *Privileged Access* that we all seem to have to the semantic and intentional properties of our own occurrent mental states. How can it be that we all seem to know in a special authoritative and non-empirical way what it is we mean when we utter certain sentences, or what we believe when we sincerely and on reflection assent to them, when what determines what we thus mean or believe pertains to external facts about which we have no special authoritative or non-empirical knowledge? This is not just an unpleasant consequence of Semantic Externalism which its proponents nevertheless may learn to live with. It alludes to deep underlying difficulties. If we did not have Privileged Access to a substantial class of semantic and mental contents, then there would be no asymmetry between the way we know of our own mental life and the way we know of others' mental life. We would, for instance, have to produce grounds provided from our own sayings and doings in order to warrant self-ascriptions of basic mental states. In a nutshell, the challenge is to accommodate the Semantic Externalist insights that semantic and mental contents are individuated, at least in part, by facts external to the individual, yet not compromise the Semantic Internalist account of how we can have Privileged Access to these contents as if they were individuated by facts internal to the individual. That sounds impossible and I think it is. More precisely, if Semantic Externalism is fuelled by Referentialism such that mental content is singular, then it is more than hard to see how we could possibly have Privileged Access to mental content given that we have no Privileged Access to the objects and properties of which singular content is constituted. Once we get clear on the semantics which facilitate the expression of such thoughts, this should come as no surprise. What Privileged Access is concerned with are only those conditions that we associate with an expression such that an object is the referent of that expression if and only

if it meets those conditions. Working under the assumption that if we have a *reductio* of *Compatibilism*, i.e. the claim that Semantic Externalism is compatible with Privileged Access, then we also have a *reductio* of Semantic Externalism, there has been no shortage of ingenious attempts to prove how the Incompatibilist Arguments fail in their intent. But hardly any Semantic Externalists have undertaken to meet the more positive challenge of actually showing in what the reconciliation could possibly consist. Even though I have no knock-down argument that the challenge could not possibly be met, I suspect the prospects look less than good and therefore that much energy has been spent in vain.

Instead I propose to discard the assumption that Semantic Externalism - the view that mental content is wide - is underpinned by Referentialism - the view that propositional content is singular. What the arguments against Descriptivism and Semantic Internalism show is not that reference could not possibly go by associated conditions, but rather how those conditions should be spelled out. In particular, those arguments highlight that not all of these conditions could be of a purely descriptive kind, but must ensure that reference is also constrained by e.g. appropriate causal connections. What I thus propound is a semantics for referring expressions which incorporates Referentialist tenets into a Descriptivist framework. It is a *hybrid view* according to which mental content is bifurcated into a narrow and a wide component. This hybrid semantics accommodates the chief virtues of both positions while also eschewing their main vices. It concedes with the Semantic Externalist that the identity of at least some of our cognitive attitudes depends on circumstances in our external, physical or social environment, but it also embraces narrow content in one sense of this notion, and so does not assign the kind of prominence to wide content that the Semantic Externalist had hoped for. In particular, it gives up on singular content and so surrenders the idea that objects and properties themselves are somehow part of content. Descriptive content with certain built-in external constraints is all the content we need. Unlike traditional Semantic Externalism, this hybrid view is not married to Referentialism, and so it paves the way for *Compatibilism*; or so I shall argue. It can both

be true that meaning is in the mind in the sense of being subject to Privileged Access, while also true that meaning is not in the mind in the sense of not necessarily being shared by doppelgängers. In short, *the kind of meaning we have in mind determines how the world determines the kind of meaning we do not have in mind*. This is the guiding thought. The mind associates with a referring expression a set of descriptive as well as non-descriptive conditions such that an object is picked out by that expression if and only if it uniquely satisfies those conditions. Knowledge of such conditions is necessary and sufficient for understanding that expression and is accessible just by introspective reflection. Nevertheless, given that some of these conditions involve reference to external circumstances, it may well happen that certain mental states have wide content. I shall proceed as follows.

In Chapter I I start off by setting up the basic problem of Incompatibilism. First of all, the case for Semantic Externalism is presented. As is familiar, this view is at least in its natural kind variant strongly supported by the so-called *Twin-Earth Argument* first advanced by Putnam. Secondly, the case for Privileged Access is presented. This phenomenon mirrors an asymmetry in psychological discourse between first-person and third-person ascriptions. A competent speaker is authoritative with respect to the contents of his own occurrent mental states. In the normal run of things, he sincerely and reflectively judges that he is in some basic mental state if and only if he is. Thirdly, a *prima facie* case for the incompatibility of Semantic Externalism and Privileged Access is discussed. A prominent proposal, due to Burge, is found to fall short of a satisfactory response to this initial difficulty. Fourthly, what has come to be known as the *MC-form*, first brought forward by McKinsey, is found to pose a serious threat to Compatibilism: I have mental property M; if I have mental property M, then I meet external condition C; so, I meet external condition C. The first premise is a priori knowable by Privileged Access, and the second premise is a conceptual truth according to Semantic Externalism, and so is a priori knowable. Hence, the conclusion is a priori knowable. But that is unbelievable given that C

embraces ordinary empirical truths, so something must give way. As Boghossian has pointed out, the second premise depends upon a kind of Semantic Externalism which is insufficiently supported by the initial Twin Earth Argument, and so a crucial distinction between Weak and Strong Semantic Externalism must be born in mind. What the Twin Earth Argument shows is that certain contents are individuated by reference to external conditions, but what the MC-form requires is that the existence of those contents depends upon the obtaining of those conditions.

In Chapter II I begin by showing the intimate connections between Semantic Internalism and Descriptivism, and Semantic Externalism and Referentialism. It is commonplace that whichever view one takes on the individuation of semantic content will have implications for the view one takes on the individuation of mental content. I then turn to the anti-Descriptivist, Modal Argument initially due to Kripke. The Twin Earth Argument presupposes that natural kind terms function in much the same way as ordinary proper names, and that these are *purely Referential* expressions, i.e. rigid designators without descriptive content. What the *Modal Argument* purports to show is that since referring expressions are modally rigid and definite descriptions modally flexible, the propositional contribution of the former cannot be identical to that of the latter. If two expressions have the same content, then by Leibniz Law, they must have the same content-individuating properties, but referring terms and definite descriptions have distinct modal properties, so their content must differ. The Modal Argument thus presupposes that modal properties are relevant content-individuating properties. Indeed it is often assumed by the Referentialist that if a referring expression is a non-empty rigid designator, then sentences containing it will only express singular propositions. The Descriptivist could, however, reject this assumption. Although prevalent intuitions tell us that referring expressions must be modally rigid, there is no straightforward entailment that they must also lack descriptive content. It is only if referring expressions are purely Referential that sentences containing them must have singular truth-conditions. So, rigidity is one thing, Referentiality another.

This paves the way for a response to the Modal Argument. Either the Descriptivist accepts the hidden premise that any two referring expressions differ in descriptive content if they differ in modal properties; in which case he must *rigidify* the definite descriptions which confer descriptive content on the referring expression. Or the Descriptivist simply rejects that premise on the grounds that descriptive content is a *cognitive* notion, individuated not by its modal properties, but by its intentional properties. The upshot is that no convincing case has been made that referring expressions lack descriptive content.

In Chapter III I present an argument - the *Intentional Argument* - originally due to Frege against Referentialism. Where the Modal Argument tried to show that two referring expressions differ in descriptive content if and only if they are not intersubstitutable *salva veritate* in modal contexts, the Intentional Argument aims to show that two referring expressions differ in descriptive content if and only if they are not intersubstitutable *salva veritate* in intentional contexts. In its strongest form, the Argument is a paradox about belief. If Referentialism is right that the meaning of an expression is exhausted by its reference, then two expressions should have the same meaning if and only if they have the same reference. That is, if ' $a = b$ ' is true, then ' a ' and ' b ' should have the same meaning, and so everywhere be intersubstitutable without change in truth-value. This, however, jars with our firm intuition that often ' S believes that a is a ' is true, while ' S believes that a is b ' is false. The Intentional Argument also contrasts with the Twin Earth Argument. Where the latter aims to show that belief can change if the world changes even though the way the world appears remains fixed, the former aims to show that belief can change if the way the world appears changes even though the world remains fixed. I discuss a response to the Intentional Argument based on considerations about conversational implicatures, and show how the Argument can be strengthened in various ways. An interesting result is that if one holds, as the Referentialist does, that the sole propositional contents of beliefs and other attitudes are singular propositions, then one is committed to ascribing contradictory beliefs even to idealised speakers *unless* those contents are rendered epistemically opaque in the

sense of not having reflective access to their logical properties. What the Intentional Argument shows, if good, is that referring expressions occurring inside the scope of an intentional operator must have truth-conditionally relevant descriptive content. But since I also hold that referring expressions occurring outside the scope of intentional operators have such content, I close the Chapter by some considerations that further support this stronger view.

After having resisted the arguments against Descriptivism and sustained the arguments against Referentialism, I conclude in Chapter IV that we should settle for *Causal Rigidified Descriptivism*. I spend some time explicating how rigidification devices work. In particular, I defend an indexical analysis to the effect that 'actually' behaves in much the same way that ordinary indexicals like 'I', 'here' and 'now' do. This analysis emphasises the two ways in which a possible world can be conceived of. We can think of a world as a way the actual world might have turned out, or given the way the actual world has turned out, we can think of a world as counterfactual. This distinction between a possible context of utterance and a circumstance of evaluation is crucial to understanding the machinery of *Two-Dimensionalism* - a recently much discussed framework that originates in Davies, Evans, Kaplan and Stalnaker. Two-Dimensionalism offers a neat illustration of the two views on offer. Speaking in the actual world, the singular content of a referring expression can be represented as a function from a world taken as counterfactual to a referent of that expression at that world. But knowledge of such a function is clearly a posteriori since one must know the relevant empirical facts about the world in which one is speaking. If understanding a referring expression requires grasp of its singular content, then understanding must therefore be an a posteriori matter. If, on the other hand, understanding requires grasp of descriptive content, then understanding comes out as a priori. The descriptive content of a referring expression can be represented as a function from a world taken as actual to a referent of that expression at that world. But since knowledge of such a function is independent of the world in which one is speaking, knowledge of descriptive

content is an a priori matter. In short, *the descriptive content of a referring expression can be represented as a function from the world of the context to the singular content expressed in that context which in turn can be represented as a function from a counterfactual possible world to a referent of that expression at that possible world in that context.* I end the Chapter by outlining how Two-Dimensionalism offers a neat explanation of the Kripkean possibilities of *a posteriori necessities* and *a priori contingencies*.

In Chapter V I return to the problems about Incompatibilism set out in Chapter 1. But first I consider the consequences of my view for the distinction between *wide* and *narrow* content. I argue that although the Twin Earth Argument has convinced us that some mental properties fail to be shared by doppelgängers, other mental properties will be narrow in an appropriate sense. My doppelgänger and I can both believe that the watery stuff is wet, but only I can believe that water is wet, and only he can believe that twin-water is wet. Water and twin-water, remember, have distinct micro-structures, but share all their surface, watery properties: being the clear, colourless liquid that falls from the sky, quenches thirst, and so on. This reflects the fact that we can share beliefs with the same purely descriptive content. Such descriptive content is *context-independent* in the sense that it is shared by doppelgängers across different contexts: one can entertain the thought that the watery stuff is wet on Earth and on Twin Earth. Descriptive content, however, is not *world-independent*: there may be far-fetched possible worlds with deviant laws of nature in which doppelgängers fail to share even purely descriptive content. But just as my doppelgänger and I cannot share certain beliefs with the same singular content, we cannot share certain beliefs with causally constrained descriptive content. I believe that the watery stuff of *my acquaintance* is wet, but my doppelgänger believes that the watery stuff of *his acquaintance* is wet. Wide content is thus context-dependent in the sense that it may fail to be shared by doppelgängers across different contexts. After having characterised narrow content as what doppelgängers within the same possible world have in common, I consider some objections to narrow content. The Semantic Externalist may try to run the Twin Earth

Argument on the watery predicates - 'liquid', 'wet', 'thirst-quenching', etc. - to show that even my thoughts about the watery stuff have wide content. In response, I suggest the assumption that these predicates have a *natural kind use* be rejected. E.g. 'being thirst-quenching' has a *superficial use* according to which, roughly speaking, any substance is thirst-quenching if and only if it plays a certain *functional role* and elicits a certain *phenomenal feel* upon consumption. The second strategy the Semantic Externalist may opt for is to show that the watery predicates are subject to Burge-style Arguments which, if cogent, apply right across the board. I suggest the same response as with the Twin Earth Argument. What these Arguments highlight is not that associated properties fail to mediate reference. On the contrary, they emphasise *which* properties mediate reference: 'water' refers to the *actual* watery stuff of *my acquaintance* which *goes under the name 'water' in my speech community*. The Social Semantic Externalist Arguments tell us that we often use expressions deferentially. The upshot is that although the Semantic Externalist Arguments have convinced us that some mental content is wide, that should not lead us into thinking that no mental content is narrow. Finally, I return to the MC-form. What one has Privileged Access to is the descriptive content commonly associated with a referring expression. It takes no empirical investigation to assure oneself that if water exists, then water is the watery stuff. Moreover, since we learn from the Twin Earth thought experiment that water is the watery stuff of our acquaintance, we can come to know just by *conceptual reflection* that water is the watery stuff of our acquaintance. So, we can know a priori that we entertain certain thoughts with wide content. Importantly, whereas both singular content and causally constrained descriptive content are wide in the weak sense of being externally individuated, only the latter is *object-independent* content. On Dry Earth, where nothing uniquely has the watery properties, utterances of sentences containing 'water' can still be assigned causally constrained descriptive truth-conditions. Not so for singular content. Utterances of such sentences do not have singular truth-conditions on Dry Earth. Thus if the Semantic Externalist sticks to Referentialism, then the second premise in the MC-form

is indeed a priori true, but the first premise comes out as a posteriori. The conjunction of Semantic Externalism and Referentialism is therefore doomed to struggle with Incompatibilism. If, on the other hand, Semantic Externalism is divorced from Referentialism, as I recommend, then the first premise is a priori knowable, but the second premise is false. Causal Rigidified Descriptivism issues in a kind of Weak Semantic Externalism according to which externally individuated, but object-independent, content is subject to Privileged Access. So, the MC-form is blocked and Compatibilism is vindicated. This, I conjecture, should be the lesson that Twin Earth has taught us.

Chapter 1. Incompatibilism

1.1. Semantic Externalism

Semantic Externalism is, in Burge's [1982, p. 149] words, the claim that the "...identity of one's mental contents, states, and events is not independent of the nature of one's physical and social environment". Or, as Putnam [1996, p. 7] put it, it is a denial of the thesis that intentional states such as beliefs and desires, "...do not presuppose the existence of anything external to the agent who possesses such states." Semantic Externalism can thus be minimally characterised as the thesis that at least certain contents of intentional states fail to supervene on intrinsic, physical properties of individuals. Accordingly, it is possible for two individuals to be exactly alike in all intrinsic, physical respects yet believe or desire different things. Externalist content, or *wide* content, as I shall call it, is content not shared by intrinsic, physical duplicates, or *doppelgangers*, as we might say. For *doppelgangers* differ only in their relational properties, in how they are related to their physical or social environment. Wide content depends instead, at least partly, for its individuation on the nature of an individual's physical or social environment in the sense that had that individual been in a different physical or social environment, she would have had beliefs and desires with a different content. And if we suppose that intentional states are individuated by their contents, the individual would also have had different beliefs and desires. Internalist content, or *narrow* content, as I shall call it, is, on the other hand, content that does supervene on intrinsic, physical properties of individuals such that no matter which physical or social environment *doppelgangers* are embedded in, they will necessarily be in the same narrow content intentional states.¹

¹ I assume that both the Semantic Internalist and the Semantic Externalist believe in psycho-physical supervenience such that no two individuals can differ mentally without there being some physical difference, whether between those individuals themselves, as the former has it, or between those individuals or their

1.2. The Twin Earth Argument

As is familiar, Semantic Externalism derives credibility from a type of thought experiment originally due to Putnam [1996]. One type of experiment goes like this: let Twin Earth be a remote planet in the actual world W_A just like Earth in every respect except there is no water, just a qualitatively indistinguishable substance with the abbreviated micro-structure XYZ. Call this substance 'twin-water'. Twin-water has all the superficial, manifest properties that water has. It is a clear, colourless, tasteless, potable liquid that fills the lakes, flows in rivers and out of taps, falls from the sky, supports life, etc. Although water and twin-water are different substances, they are both, in short, *watery* stuff. Furthermore, my doppelgänger, i.e. my internal physical duplicate, on Twin Earth is stipulated to be just as ignorant about chemistry as I am on Earth. The claim is then that our respective utterances of sentence-types containing 'water' have different semantic properties. On Earth my tokens will refer to H_2O , whereas on Twin Earth, my doppelgänger's tokens will refer to XYZ. On the assumption that meaning determines extension, so that a difference in extension suffices for a difference in meaning, my tokens will thus mean water while my doppelgänger's tokens will mean twin-water. But since we share all intrinsic mental properties - e.g. we have the same stereotypical beliefs about the surface properties of water - it follows that what we have in mind cannot determine what we mean by utterances of sentences containing 'water'.²

external environments, as the latter has it. Although it is commonplace to include physical and, at times, functional properties in the supervenience base, there is clearly an issue about whether non-physical properties could go into the base. Thus if one feels no inclination towards physicalism, then one would no doubt wish to include non-supervenient non-physical properties in the base or, alternatively, deny supervenience altogether. I shall henceforth not consider such views.

² In setting up the thought experiment it is thus assumed that doppelgängers have certain content-bearing mental states in common like the belief that something is a wet liquid. A thorough-going Semantic Externalist would insist that this assumption is more of a dialectical point than a real concession to the Semantic

Moreover the lesson that certain semantic properties are relational carries over to the corresponding intentional properties. That is, although Putnam's parable focused initially on semantic content purporting to show that ""meanings" just ain't in the *head*" [1996, p. 13], Burge [1979] was quick to point out the obvious implications for mental content. If, as we suppose, we can normally express the content of our belief that water is wet by the sentence 'water is wet', then the content of that belief will be fixed by the meaning of that sentence - the sentence which says what the belief is about. Hence, if the meaning of that sentence is individuated by external facts, then so will the content expressible by means of it. If we moreover suppose that it is partly the content of the belief that water is wet that makes it the belief it is, then if that content is individuated by external facts, then so will be the belief itself. So, not only do we have different De Re propositional attitudes - I believe of water that it is wet whereas my doppelgänger believes of twin-water that it is wet - we also have different De Dicto attitudes - I believe that water is wet whereas my doppelgänger believes that twin-water is wet. Both assumptions are at least *prima facie* plausible. Perhaps some thoughts are linguistically inexpressible - to the extent that animals have thoughts then they presumably are - but for competent speakers most if not all thoughts do seem to be in principle expressible in language. It is also plausible that beliefs are individuated, at least in part, by the truth-conditions of their content-clauses. If content is truth-conditional, then, if the contents of two beliefs are different, then the two beliefs will also be different.³ So, my doppelgänger and I will be in different wide content mental states, since the contents of our beliefs will depend upon the different causal relations that

Internalist: once it has been shown that our tokens of 'water' have different semantic properties, the same line of reasoning can be used to show that our tokens of 'wet', 'liquid', etc., also have different semantic properties. Wide content mental properties thus fail to supervene on intrinsic physical and, for the sake of argument, narrow content mental properties. I shall evaluate the scope of the Twin Earth Argument in Sec. 5.3.

³ I shall not discuss this assumption here, but simply take it for granted. It will certainly be true that if the semantics for belief ascriptions is compositional, then if two De Dicto belief ascriptions have different truth-conditional contents, then the whole ascriptions will also have different truth-conditions.

we, or our fellow-speakers, bear to our respective external environments. So, if content is wide, then the property of having an attitude with that content cannot be intrinsic. In sum, with Burge's words [1996, p. 343], the thought experiments'

"common strategy is to hold constant the history of the person's bodily motion, surface stimulations, and internal chemistry. Then, by varying the environment with which the person interacts while still holding constant the molecular effects on the person's body, one can show that some of the person's thoughts vary. The upshot is that which thoughts one has [...] is dependent on relations one bears to one's environment."

Let me illustrate. Suppose that I on Earth reach out for a cup of water to quench my thirst. I have a belief that I would express by saying 'water is thirst-quenching'. Suppose also that my doppelgänger on Twin Earth reaches out for a cup to quench his thirst. He also has a belief he would express by using the sentence-type 'water is thirst-quenching'. It is still true that chemistry is unknown to us and that we share a set of stereotypical beliefs about the superficial properties of water. Everything is the same except that I drink H_2O , whereas my doppelgänger drinks XYZ. Now, my utterance on Earth of 'water is thirst-quenching' is true iff water, i.e. H_2O , is thirst-quenching, whereas my doppelgänger's utterance is true iff twin-water, i.e. XYZ, is thirst-quenching. Our respective tokens of the same sentence-type have different truth-conditions. But if our utterances have different truth-conditions, then they must also express different contents if we think content is individuated by its truth-conditions. It will of course also be true that different De Re belief ascriptions are true of us: I believe of water, i.e. H_2O , that it is thirst-quenching, but my doppelgänger believes of twin-water, i.e. XYZ, that it is thirst-quenching. A De Re belief is trivially a relational property of an individual, because it is individuated by an object other than the individual who has it and so involves standing in an appropriate epistemic relation to this *res*. The interesting claim is that also different De Dicto belief ascriptions will be true of us - provided again that such ascriptions are individuated by the truth-conditions of

their content-clauses. My doppelgänger and I are in different De Dicto belief states: I believe that water is thirst-quenching, whereas my doppelgänger believes that twin-water is thirst-quenching. For how could my doppelgänger possibly believe that water is thirst-quenching when his environment contains no water for his belief to be about? Moreover, to the extent that ascriptions of beliefs containing embedded occurrences of 'water' were interpreted as beliefs involving the concept water, my doppelgänger's beliefs would all be false, but it would surely be more charitable to ascribe to him true beliefs involving the concept twin-water. The important point is that even such De Dicto attitudes are relational properties of individuals, because the individuating contents are wide. Hence, the Semantic Internalist's claim that all mental content supervenes on intrinsic physical properties of individuals is false. Were that supervenience claim true, two individuals exactly alike in all intrinsic respects would also be exactly alike in all mental respects, i.e. any difference in mental properties would entail a difference in intrinsic properties, but we have seen that two individuals may differ in mental properties while sharing the same intrinsic properties. We can put the Twin Earth Argument like this:

(A) Suppose Twin Earth is a remote planet in W_A just like Earth except that XYZ, and not H_2O , is the watery stuff, and suppose also that my doppelgänger on Twin Earth and I on Earth both assent to 'water is wet'.

(B) Then, my utterance of 'water is wet' is true iff H_2O is wet, whereas my doppelgänger's utterance of 'water is wet' is true iff XYZ is wet.

(C) But given that a difference in truth-conditions entails a difference in content which cannot be explained by a difference in intrinsic properties, the content of 'water is wet' is wide.

(D) If the content of 'water is wet' is wide, then my belief that water is wet is a relational property.

(E) My belief that water is wet is a relational property.⁴

Do (A)-(E) refute Semantic Internalism? Not on the face of it. Even Putnam [1996, p. 19] admitted that, strictly speaking, all we have is a disjunctive conclusion: either content is not narrow or content fails to determine extension. Having found the second disjunct unacceptable he concluded that content is wide. The thought seems to be that if content determines extension, then if extensions differ, then so must content. But extensions do differ, hence so must content. But on the assumption that the content of 'water' is given by 'the watery stuff', content does not differ, so either that assumption is false or content does not determine extension. To save the assumption, the Semantic Internalist could allude to the fact that we already have a range of referring expressions whose contents do not determine their extension in a context-independent way. Take indexical expressions like 'I', or descriptions like 'the mayor', which pick out different objects when uttered in different contexts. In support, it may be pointed out that since H₂O and XYZ equally have all the watery properties, there is also a sense in which our respective utterances coincide in truth-conditions: 'water is wet' is true iff the watery stuff is wet. Such truth-conditions are *descriptive* in that the right-hand side of the bi-conditional merely stipulates a descriptive condition under which the sentence on the left-hand side is true. In contrast, the truth-conditions given in (B) above - Earthly tokens of 'water is wet' are true iff H₂O is wet - are *singular* in that they involve mentioning the particular state of affairs under which the sentence is true. Descriptive truth-conditions are narrow as they do not supervene on our relational properties, but are contents my doppelgänger and I have in

⁴ There is a difficult issue about the *ontological implications* for belief states: does the Semantic Externalist claim that beliefs and other attitudes are not internal mean that they are external to the mind like tables and chairs are? I don't think this has to follow. The Semantic Externalist can maintain that cognitive states are located within subjects, indeed being identical to physical states of subjects, and yet admit that any adequate specification of what it is to be in such states must necessarily advert to circumstances external to the subject. What makes my beliefs about water wide is not that they are located where the water is, but that they cannot be described without reference to water.

common. There is no claim that such truth-conditions somehow fall short of constituting a coherent notion of content. The content expressed by a token of the sentence-type 'water is wet' specifies determinate conditions under which it is true or false, and the fact that in different contexts different tokens are true in virtue of different states of affairs does not show that somehow those conditions are incoherent or not fine-grained enough; or so the thought goes. The Semantic Internalist thus suggests (B) replaced with:

(B*) My utterance of 'water is wet' is true iff the watery stuff is wet, and my doppelgänger's utterance of 'water is wet' is true iff the watery stuff is wet.

There is no difference in truth-conditions, hence no difference in content, which cannot be explained by a difference in intrinsic properties. So, if our beliefs are individuated at least partially in terms of the content given by the watery predicates, then we both believe that water, i.e. the watery stuff, is wet. It is true that different De Re belief ascriptions are true of us, but nothing prevents us from sharing the narrow content De Dicto belief that water is wet. To presuppose that a difference in the kind of stuff that make our beliefs true entails a difference in the contents of our beliefs, is to presuppose that 'water is wet' does not have descriptive truth-conditions. So, the Twin Earth Argument does not pose a non-question-begging threat to Semantic Internalism.

Nevertheless, this Semantic Internalist response does seem to have counter-intuitive consequences. If 'water' simply means the watery stuff, and different kinds of stuff are watery within W_A , then not only does 'water' pick out different kinds of stuff within W_A , water *is* different kinds of stuff within W_A . With the meaning 'water' has here on Earth, it is true to say that on Earth water is H_2O and on Twin Earth XYZ is water. The Semantic Internalist may of course bite the bullet and accept these consequences. In support, it could be adduced that such a view turned out true for some natural common nouns. The extension of 'jade' comprises both jadeite and nephrite, and the extension of 'vitamin' includes a variety of distinct nutritious compounds. But widespread intuition has it that manifest natural kind terms like 'water', 'gold' or 'aluminium' are different. If water is H_2O on Earth,

then water is everywhere H_2O . In particular, water is not XYZ on Twin Earth. The stuff on Twin Earth is not water, but merely an *epistemic counterpart* of water: twin-water *appears* to be water in that it shares all the watery properties with water, but due to its distinct micro-structure twin-water *is* not water.

This claim needs support. Why is it impossible that water is XYZ given that water is H_2O ? Consider what happens when physical science uncovers so-called theoretical identifications. Take a macro-physical property like temperature. An object's temperature is functionally defined in terms of the *causal roles* it play in relation to its other properties as well as its interaction with objects in proximity: if a lump of iron is heated by a burner, then it glows, its magnitude increases, and it melts wax in its vicinity, etc. An object's temperature is the property which occupies this causal role, i.e. the property causally responsible for this cluster of laws. As a matter of fact, science discovered that in our world it is the micro-physical property of gases of having a certain mean molecular energy which uniquely fills this role. So, we conclude that temperature *is* mean molecular energy.⁵ But why not stick to the weaker claim that, as it turns out, temperature and mean molecular energy are invariably correlated? Because such identifications enable us to *explain* why certain observed macro-physical regularities hold in terms of underlying micro-physical structures given the physical laws and constitution of our world. The thought is that nothing short of micro-physical reduction would do the job. The same is presumably true of identifications between water and H_2O . We can give reductive explanations of facts about water in terms of facts about H_2O , e.g. that the boiling point of water is 100°C at 1 atmosphere is due to the molecular changes that set in under those conditions. But if a number of different micro-physical properties within the same possible world were equally good candidates to fill the causal role that water plays in our world, then no such micro-reduction could be carried out, and so no reductive explanation of observed macro-properties in terms of underlying micro-properties would be achievable. The fact that such

⁵ For an account of theoretical terms and identifications along these lines see Lewis [1970, 1999c].

theoretical reductions have been so pervasive in science - think of light and electromagnetic radiation or the gene and the DNA molecule - has, I suspect, formed our intuitions about Twin Earth cases. Twin Earth aims, as any other thought experiment, to clarify conceptual relations by appeal to more or less well-entrenched intuitions. What it purports to show is that our concept of water and our concept of the watery stuff are distinct since the correct way to describe Twin Earth is as a place where something is watery without being water. On Twin Earth where the governing laws of physics are identical to those on Earth, it is simply wrong to call XYZ 'water' in light of the fact that H_2O is the sole occupier of the causal roles that water play on Earth.⁶ XYZ share with H_2O all the watery properties, true enough, but all that means is that XYZ is an epistemic counterpart of water; XYZ is not water. What constitutes water is not its surface properties, but its micro-structure.

Consider some other possible cases. Suppose that Twin Earth is a far-off planet in our galaxy where there is a kind of stuff the inhabitants call 'gold'. It has all the manifest properties gold has on Earth: yellow, hard, precious, rare, soft if pure, used in jewellery, extracted from fossils in mines, melting point $1063^{\circ}C$, and so on. Now suppose I were to travel to Twin Earth in a space-ship and came across this stuff. Would it be correct to call it 'gold'? Surely not. Gold is all and only stuff that occupies the causal role that gold plays on Earth: Au as it turns out. The stuff on Twin Earth is an epistemic counterpart of gold, twin-gold, or better, fool's gold. Or suppose these course of events took place on Earth. By a stroke of genius, a corrupt scientist invents a chemically complicated compound which passes whatever test best science deploys to determine whether something is gold. Later the compound is distributed with big profit throughout criminal circles as 'stolen gold'. Although the compound is not found naturally, it is next to an epistemic counterpart of gold. But it is not gold. Once the fraud is revealed, it loses its market value. Or suppose

⁶ Bear in mind that XYZ was supposed to be an abbreviation for a chemical compound radically different from H_2O . It is not ruled out that isotopes of water, e.g. D_2O , or 'impure' water of some kind may fall within the extension of 'water'.

that an anthropologist discovers a civilised tribe which uses a white, fine-grained, water-soluble stuff extractable from sea-water and with the same nutritious effects that we use common salt. It tastes of salt, is called 'salt', and is in every way indistinguishable from salt yet turns out not to be NaCl at all. Although this stuff plays the functional role salt plays in our lives, it is clearly not common salt.

On this background, it is easier to appreciate what is amiss with the descriptive truth-conditions in:

(B*) My utterance of 'water is wet' is true iff the watery stuff is wet, and my doppelgänger's utterance of 'water is wet' is true iff the watery stuff is wet.

Importantly, the claim is not that only the singular truth-conditions in:

(B) My utterance of 'water is wet' is true iff H₂O is wet, and my doppelgänger's utterance of 'water is wet' is true iff XYZ is wet,

gets things right with respect to Earth. Water is H₂O because H₂O fills the watery role on Earth. The problem arises only on Twin Earth. According to (B*), my utterance of 'water is wet' is true at Twin Earth because, on Twin Earth, the watery stuff is wet.⁷ Following (B*), 'water' picks out whatever 'the watery stuff' picks out, regardless of its micro-structure. But in light of those considerations just mentioned, it is incorrect to call the watery stuff on Twin Earth 'water'. There is no water there, and so it is false that Earthly tokens of 'water is wet' are true at Twin Earth. And, the Semantic Externalist continues, only (B) can accommodate this fact. The condition under which my tokens of 'water is wet' are true at Twin Earth is that the watery stuff on Earth, i.e. H₂O, is wet. And since that condition does not obtain on Twin Earth, my tokens are false at Twin Earth.

But if Earthly and Twin Earthly tokens of sentences containing 'water' have singular truth-conditions, as in (B), then what my doppelgänger and I mean by 'water' is thus not solely a function of our shared internal make-up, but depend on unknown different facts

⁷ Let me stress that by saying that a sentence is *true at* Twin Earth I mean that the proposition it expresses on Earth is true at Twin Earth, and not that it would express a true proposition were it uttered on Twin Earth.

about our physical environments. In particular, what I mean by 'water' depends on which kind of physical stuff occupies the causal role that water plays. And if semantic content is wide, then so is mental content: I have the concept water so I can entertain thoughts involving that concept, whereas my doppelgänger who has the concept twin-water can entertain thoughts involving that concept.

This completes the Twin Earth Argument. There is a lot to say and I shall come back to the Argument in Chapter 2, 4 and 5, but this will suffice for present purposes. Let me now turn to the Privileged Access Thesis.

1.3. Privileged Access

Privileged Access is the thesis that a competent thinker S can have a priori knowledge of the intentional properties of the contents of his own thoughts. By a priori knowledge I mean knowledge arrived at introspectively, hence independently of any empirical enquiry into the thinker's external environment.⁸ S need not consult his physical or social surroundings by deploying his perceptual apparatus in order to determine the contents of his own cognitive attitudes. Other speakers, in contrast, are left with no choice but to attain adequate empirical knowledge of S's linguistic and physical behaviour if they wish to know of such contents. S knows what he thinks or feels before he speaks or acts, and even if he could invoke such behavioural evidence, there is hardly ever any reason to. S has a Privileged Access to his own psychology that other speakers lack. They always have to appeal to such evidence by means of observation or testimony to back up their third-person ascriptions.

⁸ I shall henceforth use 'a priori' to characterise the kind of knowledge involved in Privileged Access. One may prefer the label 'non-empirical' to distinguish it from a priori knowledge of non-empirical subject-matters, e.g. knowledge of mathematics and logic. It seems to me, however, that whether a piece of knowledge is a priori or not, is a question about the way in which it is arrived at, and not what it is knowledge of. Another option is to call knowledge a priori only if one's justification does not derive from 'outer' perceptual experience or from 'inner' awareness of one's mental life.

Knowledge of one's own inner goings-on is accessible in a special first-person manner, whereas knowledge of others' mental life is constrained by whatever epistemic standards govern knowledge in general. This issues in an *asymmetry* in ordinary psychological discourse between first-person and third-person utterances. The latter are characterised by avowals which are claims about, or expressions of, one's own occurrent mental or sensational states. One may distinguish between *phenomenal avowals* like 'I have a headache' or 'I feel great today' and *attitudinal avowals* like 'I hope the weather stays dry' or 'I reckon that apples are wholesome'.⁹ Both have three characteristic marks:

They are received as *Authoritative*, that is, if a speaker S understands the claim that he is F, where F is some occurrent mental state, and is sincere in making it about himself, then there is a strong *prima facie* case that S is not wrong about him being F. Indeed if S's claim is of the phenomenal kind, it is hard to see how the combination of conceptual competence and honesty could fail to guarantee the truth of his self-ascription. If S avowed 'I am tired' and there was no reason to doubt his proficiency or sincerity, it would normally be inappropriate to question the correctness of his utterance. Not so for attitudinal avowals. First of all, there are second-order beliefs in which one engages in interpretation of one's first-order attitudes and, as is well-known from the Freudian literature, in such cases one may be radically mistaken about one's own intentional life. But there are also basic attitudinal cases which do not involve self-deception or other illusions. If S sincerely and understandingly uttered 'I believe that water quenches thirst', then S has thereby provided a third-person with a reason for thinking that S presently believes that water is thirst-quenching. Of course, the truth of S's avowal is answerable to future behaviour: if S refused to drink water if thirsty and presented with water, then, other things being equal, a third-person S* would make better sense of S's overall psychology if he did not take S's avowal at face value, indeed such behaviour would be a reason not to. Perhaps something similar could happen with certain phenomenal avowals. It is conceivable that S subsequently

⁹ Cf. Wright [1991, 1998].

discovers that on a given occasion he has mistaken, say, an itch for a tickle without being insincere or conceptually incompetent. Anyway, the point is that the mere defeasibility of most, if not all, avowals provides no reason to think that the presumption in favour of their truthfulness is actually defeated. Even if both first- and third-person ascriptions are fallible, only avowals are accredited with authority. The fact that S is sincere and apprehends the relevant concepts provides no reason for thinking that his ascription of a phenomenal or attitudinal state to someone else is true, unless he can produce some evidence originating from that speaker's behaviour. For an observer S* to have a reason to believe that S is sincere and understands his avowals is *eo ipso* to have a reason to believe S's avowals, but S*'s competence and sincerity is no reason to believe her corresponding third-personal claim, absent additional information concerning S's sayings and doings.

The second characteristic concerns the *Non-inferential* nature of avowals. A speaker S may justifiably assert that he is in some basic mental state F without being required to adduce grounds to support his assertion. Unlike third-person ascriptions, it would ordinarily be misplaced to demand of S evidence in support of his claim that he is F. Suppose S sincerely and competently avows 'I've a sore throat' in the presence of an observer S* who later reports to another speaker S** 'S has a sore throat'. Then suppose S** replies: 'How do you know?' We are inclined to think that this question is appropriate only if addressed to S*. S knows directly by introspection how things presently are with him and that is why there is no need for him to back up his self-ascriptions of basic mental states. S does *not infer* from his own verbal and physical behaviour that he is F. Even in cases in which such evidence is available, he does not consult it - he need only reflectively attend to his state F. Other speakers, however, have no such immediate access to S's basic mental states, but must rely on his sayings and doings as their inferential justifying basis.

The third and last distinctive mark is the *Transparency* of the objects of avowals. A basic mental state F is *salient* to the speaker who has it. This means that if, in normal circumstances, F occurs in S, then it would seem to be necessary that S knows that F

occurs. In similar circumstances, if *F* does not occur, then it cannot seem to *S* in every way as if *F* occurs. Although not infallible, *S*'s 'inner sense' is not hostage to the kind of perceptual illusions that afflict his 'outer sense', since its subject-matter does not exhibit the same kind of appearance-reality distinction. Unless *S* suffered from some cognitive or other disabling malfunction, it would be very odd for a basic state like a headache or a desire for chocolate to occur in *S* without *S* being aware of its presence and disposed to give the thought that he is in such a state expression. In the normal run of things, it is hard to see in what *S*'s ignorance of the truth-values of his own self-ascriptions could consist. Not so for other speakers. It is possible for *S** to know all relevant facts about *S*'s recent behaviour yet be ignorant of features of his basic psychology. *S*'s beliefs and desires are not salient to anyone but *S*.

The foregoing features - Authority, Non-inferentiality and Transparency - should be commonplace. There is, however, also a stronger and less uncontroversial sense in which *S*'s phenomenal and intentional mental contents are transparently available to him. According to this line of reasoning, it is also a consequence of Privileged Access to such contents that a competent speaker *S* knows introspectively, hence a priori, whether any two intentional contents towards which he has a cognitive attitude are the same or different. Why is that? Well, for *S* to have a priori knowledge of an occurrent belief is not merely to *know that* he has a certain belief - it is also to *know which* belief that is. It involves a priori knowledge of what it is to have a belief with the content *p*, e.g. how that belief is inferentially related to other beliefs, how having it disposes him to behave in various ways given various desires and physical stimuli, etc. But to have such knowledge must presuppose an ability to introspectively discriminate between *p* and other contents. If that were not so, it would be impossible for *S* to reflectively determine the inferential and behavioural consequences of his belief that *p*. If a speaker *S* thus believes that *p* and also believes that *p**, and *p* and *p** are the same mental contents, then *S* must know a priori that they are the same contents, and so must know a priori that his belief that *p* and his belief

that p^* are identical, given our assumption about individuation of beliefs by their contents. Conversely, if S believes that p and also believes that q, and p and q are different mental contents, then S must know a priori that they are different contents, and so must know a priori that his beliefs are different. In this sense, mental content is *Epistemically Transparent*. This idea was captured by Dummett [1978, p. 131], cf. also [1973, p. 95], [1991, 126]:

"It is an undeniable feature of the notion of meaning - obscure as that notion is - that meaning is *transparent* in the sense that, if someone attaches a meaning to each of two words, he must know whether these meanings are the same."

Although Dummett speaks of meaning and not mental content, the point is the same. Had meaning not been transparent in this sense, it would be possible for S to believe that p and to believe that p^* , yet profess ignorance as to whether he believes the same or different things. But since we are inclined to think that S must know a priori whether he believes the same or different things, if he is fully attentive and conceptually competent, it does seem plausible that meaning should be Epistemically Transparent; or so the thought goes. It is, to repeat, important not to lump these two senses of transparency together. It is clearly possible to hold that content p is transparent in the weak sense that, very roughly, if p occurs, then S believes that p occurs, without it being transparent in the stronger sense that S is able to tell whether p is identical to, or distinct from, some other content q that S grasps.

A speaker S thus standardly credits himself with a *Non-inferential Authority* as regards the *Transparent* contents of his own mind, and, importantly, this authority issues in a priori knowledge of those contents. In contrast, a third-person S^* cannot justifiably assert that S is in a mental state without producing grounds provided by inference from S's linguistic or physical behaviour. Roughly speaking, we may sum up *Authority* by saying that if S judges that S has p, then S has p, and *Transparency* by saying that if S has p, then

S judges that S has p, where p is some occurrent mental state. So, S has p iff S judges that S has p. Clearly, as formulated this bi-conditional calls for *ceteris paribus* clauses given the various ways in which knowledge of our own mind is defeasible. But if we could write down the relevant clauses - clauses ruling out self-deception, inattention, incompetence and what not - then it should come out a priori. It is no empirical discovery that the epistemology of mental self- and other-ascriptions reflect asymmetries between the way I know my own mental life and the way I know others' mental life. As regards the latter, a corresponding bi-conditional is not even close to being true. In any case, it is illustrative of what is to come.¹⁰

1.4. Incompatibilism

The foregoing provides sufficient background for setting up at least a *prima facie* case against Compatibilism, i.e. the view that Semantic Externalism and Privileged Access are consistent doctrines. It is easy to see the worry. How can we be authoritative about the contents of our own minds if, as Semantic Externalism tells us, those contents depend for their individuation on external circumstances that we have no special authority about, or may even lack knowledge of? In other words, how can we have Privileged Access to what

¹⁰ A different and a very deep philosophical problem is how to explain what makes it the case that avowals exhibit Privileged Access, i.e. what it is in virtue of which each one of us enjoys this kind of privilege. I shall not try to sketch such an account here; it suffices for my purposes to note the existence of an asymmetry. The most promising account that I know of is Wright's response-dependence analysis: S's psychological self-ascriptions are in the best cases constitutive of their subject-matter; see his [1991, 1998]. An alternative approach is to opt for projectivism: avowals are not truth-apt assertions that purport to represent inner reality, but simply expressions of one's psychology. The difficulty that any satisfactory analysis must face is to find middle ground between the Cartesian who views the truth-conferring mental states of affairs as occurrences in a private arena exclusively accessible to the speaker and constitutively unrelated to any outer goings-on, and the Rylean who, in reaction to the misconceived Cartesian account, overplays the role that public behaviour plays in constituting mental states, and thereby jeopardises not just our intuition that at least phenomenal states like pains are genuine mental occurrences, but also the very thesis of Privileged Access.

our thoughts are about if what makes it true that our thoughts are about what they are about resides in external facts that we have no Privileged Access to? Take our Twin Earth story. The fact that on Earth my tokens of sentences containing 'water' express water thoughts is partially due to certain causal relations that I bear to my physical environment on Earth. Had I been on Twin Earth, my tokens of the same type of sentences would have expressed twin-water thoughts due to the different causal relations I would have borne to my physical environment on Twin Earth. What goes on inside me thus falls short of fully fixing such thoughts. Yet it is what goes on inside me that I can access just by introspective reflection; the fact that I sustain particular causal relations does not seem reflectively available. So, the worry is how I could fully know just by introspection which wide content thoughts I am entertaining? It is all the same to me from the inside, as it were, whether I think water thoughts or twin-water thoughts. Would I not need to consult my external environment? Could someone who knew the relevant external facts not be in a better position than I am to know the wide contents of my thoughts?

It is important not to confuse this *prima facie* worry with this line of argument: (i) in order to know a priori that I think water thoughts, I have to know a priori that I do not think twin-water thoughts, since had I thought twin-water thoughts, I would not have thought water thoughts; (ii) to know a priori that I do not think twin-water thoughts is to know a priori that I am not on Twin Earth; (iii) but I cannot know a priori that I am not on Twin Earth; so, by *modus tollens* on (ii) and (iii), (iv) I cannot know a priori that I do not think twin-water thoughts; so, by *modus tollens* on (i) and (iii), (v) I cannot know a priori that I think water thoughts. (i) is flawed. Compare with (i*) In order to know that I have two hands I have to know that I am not a brain-in-a-vat, since had I been a-brain-in-a-vat I would not have had two hands. By substitution throughout we get that I cannot know I have two hands. (i) and (i*) presuppose that in order to know a proposition p, one must know any proposition q whose falsity is inconsistent with the truth of p. But such a - Cartesian - principle is highly dubious. In many cases, I can know p whether or not I know q as long as

q is in fact true. Thus I can know, just by looking, that I have 20 pence in my hand although I am ignorant of the causal conditions that enable my perception and cannot rule out the possibility of counterfeit coins. Presumably, for each p the number of inconsistent false qs is very high and it seems that I am standardly entitled to disregard them unless a specific reason is provided why that would be illegitimate - unknown to me a large number of fake coins have just been put into circulation. In any case, even if such a principle could be vindicated, the Incompatibilist would have to rely on an explicit sceptical line of reasoning which did not initially motivate her concern.

The worry is thus not that one has to know all the individuating conditions q of mental content p prior to one's knowledge that p, but how one could as much as know p a priori in the first place given that those qs are a posteriori knowable. This *prima facie* tension between Semantic Externalism and Privileged Access is widely acknowledged, even by Compatibilists like Burge [1996, p. 343]¹¹:

"Our problem is that of understanding how we can know some of our mental events in a direct, non-empirical manner, when those events depend for their identities on our relations to the environment. A person need not investigate the environment to know what his thoughts are. A person does have to investigate the environment to know what the environment is like. Does this not indicate that the mental events are what they are independently of the environment?"

Nevertheless, two points are often put forward to show that the appearance of tension is illusory. Consider this version of our proximate bi-conditional: S believes that he believes that p iff he believes that p. It follows from left-to-right that S cannot believe he has a belief he does not have, and from right-to-left that S cannot believe he does not have a belief he does have. If S believes he has a belief, then he has it, and if S has a belief, then he believes

¹¹ See also Davidson [1996, p. 330].

he has it.¹² What the first point - Burge's point¹³ - says is that this bi-conditional holds trivially in certain *basic* cases. Suppose:

(1) S is thinking that water is thirst-quenching,

is true. By 'thinking' is meant entertaining the thought expressed by 'water is thirst-quenching'. In order to have that thought S must at least partially understand the sentence that expresses it. Now suppose:

(2) S judges that (1),

is also true, i.e. S accepts the thought expressed by (1) as true. But to do that S must engage in the thought expressed by 'water is thirst-quenching'. So, if S judges, or merely entertains the thought, that S is currently thinking that water is thirst-quenching, then S is currently thinking that water is thirst-quenching. (1) is true if (2) is, no matter which empirical evidence S has or lacks. (2) is akin to Descartes' *Cogito*-thought *I am thinking* in that both are self-verifying. *Cogito*-like thoughts are self-verifying judgements about thoughts in the sense that the mere judging guarantees the truth of what is being judged. There is no need to support the judgement with empirical evidence even if available. In other words, in thinking knowledgeably that one is thinking that P, one is *eo ipso* also thinking that P. In such episodic cases there can be no error based on a gap between the first-order thought and the second-order thought - no possibility of misfit - because the first-order thought is simply contained as the object of the second-order thought.¹⁴ Burge's point is thus fairly

¹² Again, as stated this bi-conditional is clearly defeasible both ways and so should be taken with a grain of salt. For instance, it should only embrace occurrent belief states: I have standing beliefs I do not occurrently believe I have.

¹³ Cf. Burge [1996, 1998].

¹⁴ Burge [1996, 1998, p. 251] calls such errors 'brute'. A perceptual judgement may be wrong in ways that has nothing to do with observational conditions, rational failures, cognitive malfunctions, etc., but simply because the objects of perception are independent of our perceptual judgements in a way that the objects of our reflexive judgements about our occurrent thoughts are not. Look-alike feet are epistemically possible, look-alike present-tense thoughts are not. The fact that second-order judgements about thoughts are self-

restricted. It is only in the very act of thinking that one is thinking that P that one can know that one is thinking that P. But this quasi-logical point about self-verifying judgements about thoughts only seems to account for a minor and rather insignificant portion of knowledge of our mental life. Reflect first that Privileged Access never was meant to be a commitment to infallibility or incorrigibility in the domain of first-person psychological utterances. Not only do mental states now and then have properties I am unaware of, it also happens that I judge them to have properties they, in fact, lack. Moreover, although less likely, mental states can occur without my awareness, and I can judge them to occur when, as it happens, they do not. But if Privileged Access is generally accounted for by the model of self-verification, then how do we explain these possible mismatches between judgement and what is being judged? On this model, judgement about mental states should never err. More precisely, only judgements of *occurrent thoughts* are self-verifying. Judgements of *standing* mental states or *occurrent events* are not. S's judgement that S believes that Scots are mean is not self-verifying since S can entertain the thought that all Scots are mean without believing it. Nor is S's judgement that S has a red afterimage self-verifying. Such non-content bearing events are not guaranteed to occur if one judges that they do. The reason why it is possible for S to judge falsely that he is in some standing belief state or that some mental event has occurred, is that the mere judging fails to make it true that he is in that state or that that event has occurred. But since we saw that Privileged Access extends to such non-*Cogito*-like states and events as well, an alternative model to that of self-verification must be found. Moreover, a second-order judgement is self-verifying only if the time at which it occurs is coincident with the time at which the first-order mental episode occurs. In no other case is the object of judgement self-referentially contained in the judgement. No self-verifying thought about what S was thinking yesterday is available to him now. Whenever S's first-order thought occurs prior to his second-order judgement,

verifying implies that there is a necessary relation between the judging and its subject-matter that has no correlate in perceptual judgements.

the latter does not contain the former as its subject-matter. But cases in which S knowingly attends to his thoughts after having entertained them do seem essential to Privileged Access. So, the model of self-verification not only cannot accommodate a range a mental phenomena that we intuitively think fall within the domain of Privileged Access, it also offers no explanation of how we can know our own thoughts immediately after we have had them. That is, we have Privileged Access to first-order thoughts that are not part of second-order thoughts, indeed most of our empirical first-order thinking is not accompanied by second-order thinking, since we are not always self-reflectively thinking about our thoughts in the very act of thinking them. In most cases we attend only subsequently to our empirical thoughts.¹⁵

A different, although closely related, point - Davidson's point¹⁶ - concerns the self-referential nature, not of the act of occurrent thinking, but of the content of such an act. Where Burge's point is that the mere *judging* that one is thinking that P guarantees the truth of what one is judging, Davidson's point is that the *content* of one's thought that P - that P - automatically carries over to one's judgement that one is thinking that P. The content of the thought in (2) is self-referentially fixed by the content of the thought in (1). It is thus not possible that S judges that he is thinking that water is thirst-quenching when, as a matter of fact, he is thinking that twin-water is thirst-quenching. The second-order thought simply contains the content of the first-order thought as its subject-matter. This point is more general than Burge's. It can embrace different kinds of propositional attitudes and allow for fallibility. If I believe that I desire a glass of water, then maybe I am wrong. I need not desire a glass of water in order to believe that I do - suppose I really desire a jar of water. But I could not be wrong for the reason that, in actual fact, I desire a glass of twin-water. So, my second-order judgement may be wrong for various reasons except that I have misidentified the externally individuated content of my first-order attitude. Whatever

¹⁵ For more discussion see Boghossian [1989].

¹⁶ Cf. Davidson [1996].

external individuation conditions hold for my first-order attitude will also hold for my second-order attitude. The possibility of having the second-order belief is thus partially grounded in the possibility of having the first-order belief. So, the present point is that external individuation does not by itself give rise to *additional* difficulties in explaining how we can have knowledge of our own minds.

Given what was said in response to the (i)-(v) argument, maybe we can now rescue Privileged Access. Note first that part of the individuation conditions for thinking the *first-order* thought in (1) are external to S, and therefore only ascertainable by empirical means. But this does not mean that S needs to know that those external conditions obtain in order to think the thought in (1), as long as they in fact *do* obtain. If S thinks that water is thirst-quenching, then S sustains adequate causal connections with water, but S need not know that he so interacts. The enabling conditions need only be presupposed, not known to be satisfied. The same applies at *second-order*: S's thinking knowledgeably that S is thinking that water is thirst-quenching is grounded in an ability to think that water is thirst-quenching, which in turn, is conditioned on the possibility of physical interaction with water. The reason for this is, as we have already seen, that the second-order thought comprises the object of the first-order thought, and not merely the first-order thought as its object. Hence, the same enabling conditions that must be presupposed, but not necessarily known to be satisfied, at first-order thinking must equally be presupposed, but not necessarily known to be satisfied, at second-order knowledgeable thinking. The upshot is supposedly that we can have knowledge of thoughts without knowing the relevant facts about the individuation conditions of those thoughts.

Does this meet our *prima facie* worry? If good, this line of reasoning shows that it is feasible to have a priori knowledge of occurrent wide content thoughts. It thus proves that Incompatibilism better not be construed as the claim that Semantic Externalism and Privileged Access are *inconsistent*. But consistency is one thing, full-fledged compatibility another. And it is certainly doubtful whether all cases of Privileged Access can be

accounted for in the manner just sketched. Suppose S were unwittingly to undergo a series of switches between Earth and Twin Earth sufficiently slow to acquire the concepts appropriate to each place. Imagine a future in which space travels by the speed of light between the planets in our solar system are commonplace. How long exactly S would have to stay at each place to make for the relevant conceptual changes is unclear. All we know is that were he slowly enough shifted back and forth, he would be able to sustain sufficient causal links with his environments to effect changes in his wide content thoughts. So, suppose S is on Twin Earth at t_1 long enough to acquire twin-water thoughts, then on Earth at t_2 long enough to acquire water thoughts, and so on. The question is: could S tell just by introspection which thoughts he entertains when the two places and the switches between them are subjectively indistinguishable? One might think the answer is negative for how could S introspectively discriminate between thoughts - water thoughts and twin-water thoughts - which from the inside, as it were, seem identical. It looks as if nothing in S's understanding of the contents of those thoughts would enable him to tell them apart. But this presupposes that S at t_2 is able to entertain both sets of thoughts and so at t_2 retains the twin-water thoughts acquired at t_1 . If all of S's thoughts at t_2 , including the ones acquired at t_1 , are water thoughts, then it is no wonder that S cannot explicate the difference between water thoughts and twin-water thoughts. S cannot discriminate between two sets of thoughts if he can only grasp one of them. The problem with this account of slow switching is that S's beliefs about his beliefs at t_1 come out false. Suppose that at t_1 S forms the true belief that he drinks twin-water. At t_2 S tries to recover his belief - a belief he would express with 'I drank water', but if what S now remembers is that he believed that he drank water, then his belief is false. So, S better retain his twin-water thoughts at t_2 . But if that is how we should understand slow switching, then it is not clear how Davidson's point is helpful in cases involving memory of past thoughts. Suppose S* knew of S's predicament and asked S: 'Did you entertain water thoughts or twin-water thoughts at t_1 ? ' Having both sets of thoughts available yet no introspective means by which he could discriminate

between them, it is hard to see how he could answer the question. The worry is thus that at least Epistemic Transparency (cf. Sec. 1.3) is compromised: the difference between water thoughts and twin-water thoughts is epistemically opaque to S in that S could have both sets of thoughts and yet lack introspective knowledge of their difference. Another worry relates to Authority: S would be in no epistemically advantageous position with respect to this aspect of his thought-activity, indeed S* who had knowledge of S's recent transportations would be more suitable to ask than S himself.

I shall not adjudicate this dispute. Maybe the Semantic Externalist has the resources to accommodate slow switching. Three things are worth keeping in mind. First of all, there is no inference from the *mere* fact that water thoughts and twin-water thoughts are subjectively indistinguishable to the claim that water thoughts are not knowable unless the environmental differences that individuate them are brought out empirically. One must show that both sets of thoughts are simultaneously accessible, and one must provide reasons for thinking that the uncongenial twin-water thoughts pose a *relevant alternative* that needs to be ruled out; whence the travelling story. Only then can slow switching be used to get the (i)-(v) argument up and running: if S, quite unawares, underwent a series of such transportations and was now thinking water thoughts, then he would have to rule out the relevant alternative that he was not now thinking twin-water thoughts. Hence, he would have to know what he could not possibly know, namely that he was not now on Twin Earth.¹⁷ Secondly, even if slow switching were to found a successful Incompatibilist argument, it may be that nothing intuitively valuable would be lost. If all slow switching would show is that Epistemic Transparency is false, then maybe that is for the best. Maybe Privileged Access should encompass only the contents of our contemporary mental states and not the identity conditions of the contents of states had over time. Thirdly, it is not

¹⁷ I shall not pursue this line any further. For elaboration see Boghossian [1989, 1994], and response in Burge [1998]. The issue depends on exactly which type of 'relevant alternative' epistemology the Semantic Externalist is committed to. For a recent development of the Switching Argument see Goldberg [1999, 2000], and Brueckner's response in [1999, 2000].

generally true that one cannot have beliefs involving a content one has an incomplete understanding of: I can falsely believe that I have arthritis in my thigh. So, even though S is in the rather awkward situation that he can know, in a special self-referential way, that he thinks water thoughts when he is on Earth, and can also know that he thinks twin-water thoughts when he is on Twin Earth, yet cannot introspectively discriminate between them, these considerations do not undermine the quasi-logical nature of thinking knowledgeably about thoughts. S can know his thoughts without knowing much about them.

Now, someone might object that my criticism is misplaced. Neither Burge's nor Davidson's point was supposed to provide a uniform model of all cases of Privileged Access. I agree and I also think their points are fine as far as they go. All I claim is that Compatibilism cannot be vindicated by adopting and extending such models, and so the *prima facie* difficulty still stands.

1.5. The MC-Form

It is, however, worth trying to bring out this apparent tension in a more cogent fashion. The most promising way of doing that is a *reductio* - due to McKinsey [1991] - of the following MC-form¹⁸:

- (3) I have mental property M,
- (4) If I have mental property M, then I meet condition C,
- (5) So, I meet condition C,

where C is an 'external proposition' that makes M a wide content mental state. For instance, if M is the state I am in when I think that water is wet, then C might be the proposition that I have causally interacted with instances of the natural kind water. In general, the Semantic Externalist could endorse the following constraint on concept-possession:

¹⁸ Cf. Boghossian [1997], McLaughlin & Tye [1998] and Davies [1998].

If the concept of X is a natural kind concept, then one cannot possess it unless one has causally interacted with instances of X.

Or perhaps it is preferable not to require actual encounters by each and every possessor of the concept. Maybe it would suffice that one be a member of a linguistic community in which other members have causally interacted with such-and-such instances. The following is less demanding:

If the concept of X is a natural kind concept, then one cannot possess it unless one has causally interacted with instances of X or one is a member of a linguistic community in which other members have causally interacted with instances of X.

In any case, both proposals have it that instances of X exist in the person's environment to whom the wide concept is ascribed. Now, suppose the mental property is possession of the concept water and that water is a natural kind concept. Then we get the following instance of the MC-form:

(3*) I have the concept water

(4*) If I have the concept water, then water exists.

(5*) Water exists.

The line of reasoning seems intuitively valid and both premises warranted. (3)/(3*) is a priori knowable by Privileged Access. (4)/(4*) is a conceptual truth according to Semantic Externalism, so a priori knowable. Hence, (5)/(5*) can be deduced on purely a priori grounds and is therefore itself a priori knowable. Yet this is incredible since C embraces statements about the existence of ordinary, external objects about which, intuitively at least, I have no a priori access. So, provided the inference is cogent¹⁹, one or both of the warrants cannot be a priori.

¹⁹ Martin Davies [1998, 2000], has argued that the MC-form illicitly trades on a subtle question-begging line of reasoning. Similarly, Crispin Wright [2000] has argued that the warrant for the first premise fails to transmit across the entailment to the conclusion on the grounds that we must already presuppose entitlement to accept the conclusion. To deal satisfactorily with their epistemological diagnosis would take me too far

To get a successful run of the MC-form, it is crucial that the conditional in (4)/(4*) be not only necessary, but conceptually necessary. Since we know that knowledge of metaphysical necessities can be a posteriori, it would be possible to block the *reductio* if the major premise could be shown to be at best a posteriori warranted.²⁰ No a priori warrant transmits across a necessary a posteriori warranted conditional. The problem with this response to the MC-form is twofold. First of all, if the conditional were metaphysically necessary, then the mental states in question would be dependent upon irrelevant facts. For instance, my belief that water is wet metaphysically necessitates the (timeless) existence of my mother, but the reason why the content of that belief-state is wide has nothing to do with my biological origin. In contrast to conceptual necessity, metaphysical necessity cannot rule out irrelevant external dependencies.²¹ Secondly, the dependency between the a priori knowable contents of our mental states and their external individuation conditions is established by thought experiments or other conceptual reasoning carried out from the philosophical armchair, so to speak, and is therefore best seen as a priori, yielding conceptual truths. Semantic Externalism and its claim about content individuation is not an empirical thesis.

Another objection that one may want to raise is to question the incredibility of the conclusion that (5)/(5*) is a priori knowable, i.e. to endorse the a priori status of certain empirical existence-claims. It is clear that the MC-form has no bite if nothing is genuinely reduced *ad absurdum*. That there can be a class of a priori knowable empirical statements is shown not only by uninformative examples like 'I exist', but also by the very mental

away from the semantic considerations I wish to highlight, so I shall simply take the reasoning for granted in the following.

²⁰ Thus Brueckner [1998] has argued that the MC-form misfires, hence that Compatibilism can be vindicated, if the major premise is seen as an a posteriori metaphysical necessity. For a response, see McKinsey [1991].

²¹ The reason for this is that nothing conceptually necessitates anything except under certain descriptions. My mother's (timeless) existence, i.e. Anna's existence, is in itself metaphysically necessitated by my existence, but her existence is conceptually necessitated by my existence only under the description 'my mother'. In this way it does not follow that Anna exists if I do, but it does follow that my mother exists if I do.

contents that we standardly credit ourselves with in Privileged Access: I can know a priori that I now entertain the proposition water is wet even though I might not have done just that. So, maybe we should embrace certain informative aspects of the outer world as well.²² The problem with this line is clearly its lack of intuitive appeal. This is best appreciated when the range of conditions E is taken into account. Bear in mind that we can run the argument on any non-theoretical natural kind term such as 'water', 'gold', 'aluminium', 'mackerel', 'tiger' or 'elm'. And if *social* Semantic Externalism²³ is allowed for, one may include terms for artificial kinds such as 'sofa' or 'football'. Finally, we can also construct Incompatibilist arguments with conclusions about our internal architecture, where the conditions E will consist in a specification of the appropriate internal, cognitive make-up of the subject to whom the state in question is ascribed. It is simply not plausible that we can carry out chemical studies or cognitive science from the philosophical armchair!

1.6. Weak or Strong Semantic Externalism

A more serious objection is that (4)/(4*) seems to be insufficiently supported by the Twin Earth Argument. As the contra-positives of (4)/(4*) make clear, it is not enough to show that certain contents of intentional states depend for their *individuation* on the nature of their physical environment. What must be shown is, moreover, that such contents depend

²² This bullet-biting strategy has been adopted by Sawyer [1998]. It is also implicit in Putnam's [1981] argument that if I were a brain-in-a-vat, then I could not have thoughts about brains-in-vats, but I do have thoughts about brains-in-vats, so I am not a brain in a vat. The combination of philosophical theorising and reflective introspection yields empirical knowledge that I am not a brain-in-a-vat. For discussion see also Davies [1998, 2000].

²³ This extension of Putnam's *natural kind* Semantic Externalism is familiar from Burge [1979, 1982]. The thought experiments are the same. One fixes all descriptions of the physical, behavioural, phenomenistic and functional histories of individuals. Then, by varying either micro-physical facts or linguistic use-facts in their respective external environments, one can change the contents of their propositional attitudes. See Chapter 5 for discussion.

for their *existence* on the existence of appropriate objects in that physical environment. And clearly all the Twin Earth Argument shows, if good, is that in cases where a natural kind concept has an extension, it is individuated externally in terms of that extension. It teaches us that having a natural kind concept is a property that fails to supervene upon intrinsic physical properties of thinkers, that is, otherwise identical thinkers may fail to share a natural kind concept if embedded in relevantly different physical environments. But the Argument is silent about how those concepts are individuated in circumstances in which the relevant physical facts go missing - it simply does not speak to the issue of whether natural kind concepts could be had were there no natural kinds to fall under them. In order to sustain the a priority of (4)/(4*), however, it has to be shown a priori that the very existence of such concepts is dependent upon the existence of an extension, i.e. that they are *object-dependent*.²⁴

Paul Boghossian [1997, 1998] has shown that such additional argument is forthcoming once we consider empty cases: imagine Dry Earth as a remote planet in W_A where, despite all appearances, the lakes, rivers, taps, etc. all run bone dry. The claim is that, following Semantic Externalism, on Dry Earth where massive reference-failure

²⁴ It is important to emphasise the difference between what I call 'object-dependence' and what Evans [1982] and McDowell [1984, 1986] have referred to under the same name. As is familiar, they held a 'no referent - no thought' view with respect to so-called De Re senses, i.e. essentially Fregean senses - modes of presentation - that could not exist in the absence of a referent of the expression with which they are associated. What they had in mind, at least in the case of perceptual demonstrative thought, was a type of thought of which *no tokens* could exist were it not for the existence in the immediate environment of appropriate contextually-determined objects. What I call 'object-dependence' is, however, a type of thought of which only *some tokens* need have that kind of dependency, but of which all tokens depend for their existence on the existence of instances of the appropriate natural kind in the global environment. Consequently, what Evans and McDowell had in mind, as I read them, was a type of object-dependency that allowed for local cognitive illusions of content, whereas the type of object-dependency I have in mind is consistent with the possibility of a thought token surviving such local illusions; on my view, it is possible to have an object-dependent thought as long as one does not inhabit Dry Earth where global hallucinations prevail.

prevails, there could be no determinate conditions under which tokens of 'water'-sentences are true or false, and so no concept of water to be had. For what concept could be expressed by tokens of 'water' on this Dry Earth? Two options appear. Maybe such tokens express (1) a *compound* concept like the concept of *the watery stuff*, or maybe such tokens express (2) *an atomic concept*. The worry is that neither will do given the commitments of Semantic Externalism. Suppose that (1) is true. Then bear in mind that Earthly and Twin Earthly tokens of 'water' did not express such a compound concept as that would have implied what the story was supposed to dispute, namely that those tokens were co-extensional. It was presupposed in non-empty cases that 'water' expresses an atomic concept. But then it is difficult to see how one and the same word can express an atomic concept under conditions of successful reference and a different, compositional concept under conditions of unsuccessful reference. The *compositionality* of concepts cannot be a function of external circumstances in this way, but must be an a priori fact about the internal syntax of a concept. Suppose instead that (2) is true. One should now ask: what proposition, i.e. truth-condition, is expressed by such empty tokens of 'water is wet' as uttered on Dry Earth. One can begin by ruling out the candidate that it should be the atomic concept water as expressed by non-empty tokens of 'water' that is expressed by empty tokens of 'water' for that would simply contradict the thesis that such atomic concepts are externally individuated. Yet no other candidate could fare better, since on the Semantic Externalist account, when there is no referent of 'water', there is no fact of the matter what proposition is expressed by 'water is wet'. So, even if such sentences are false in all possible worlds, nothing really gets said by those empty tokens. Boghossian [1998, p. 255] concludes:

"If there is no kind denoted by a given natural kind term - say, 'water' - then no satisfaction conditions for that term will have been pinned down. The actual kind of stuff at the end of the relevant causal chain is supposed to fix the kind of stuff that is denoted by 'water'. If there's nothing there, then, it would seem, there is no fact of the matter what kind of stuff there would have to be for the extension of the term not to be empty."

The train of thought behind Boghossian's argument thus seems to be that if one could have a wide content belief without the existence of appropriate external objects, then that belief cannot essentially be individuated by reference to those objects, since were that belief thus essentially individuated, reference would have to be made to those objects in order to have that belief. In short, if understanding is indifferent to the existence of water, then it is also indifferent to the substitution of water for twin-water. But since it is not indifferent to the substitution of water for twin-water, neither is it indifferent to the existence of water. The upshot is that even if the Twin Earth Argument only licenses external individuation, the Twin Earth Argument in conjunction with the Dry Earth Argument entails object-dependence. This is obviously a *Strong* form of Semantic Externalism: if there is no fact of the matter what the referent of a given concept is, then there is no fact of the matter what the concept is. If this is right, then natural kind concepts are object-dependent.²⁵ The argument is thus that merely Weak Semantic Externalism is an unstable position that invariably will collapse into its stronger counterpart.

Now, I suspect the most intuitive response to the Dry Earth challenge is to opt for (1): what is said by use of 'water' on Dry Earth is that there is a stuff that uniquely has the watery properties. Although nothing falls under it, this - default - concept can be expressed even on Dry Earth. The problem is how the Semantic Externalist could meet those intuitions. Reflect that what supposedly happens on Dry Earth with respect to water

²⁵ Cf. Boghossian [1997, p. 174]. This means that on the Semantic Externalist account, occurrences of natural kind terms in the embedded sentences in De Dicto belief ascriptions allow for existential generalisation. If I have the wide content De Dicto belief that water is thirst-quenching, then I also have a corresponding De Re belief involving water, i.e. there is something - water - of which I believe that it is thirst-quenching. Indeed it is only in virtue of having the De Re belief that I can have the De Dicto belief. In contrast, a narrow content belief is given by a De Dicto belief ascription whose embedded sentence contains no expression which allows for existential generalisation. This follows from the quantificational treatment of 'the watery stuff': to believe that the watery stuff is thirst-quenching is not necessarily to believe of anything that it is watery. One can have this belief on Earth, Twin Earth or Dry Earth. I shall return to this point later.

actually happened on Earth with respect to Phlogiston. For a while it was common opinion that phlogiston was a natural kind, but it turned out not to be - in fact and despite appearances, it turned out not to exist. Had phlogiston existed as a genuine natural kind, and had 'phlogiston' aimed at naming that kind, our concept of phlogiston would have been externally individuated. In actual fact, it turned out not to exist, and the question is how best to make sense of speakers' linguistic behaviour involving 'phlogiston'. My hunch is that this is best done if 'phlogiston' is associated with a compound concept of the kind of substance that is present in all combustible materials, is released during burning, etc., since such a concept is not object-dependent. In case nothing uniquely has all the superficial properties, there are still determinate conditions under which sentences containing 'phlogiston' are true or false. So, there is no question a compound concept can be retained after the discovery that the natural kind it aimed at picking out was not in fact part of the fabric of the world. In other words, what was discovered was not that we had all been victims of a massive illusion of content, but that the world had let us down by failing to provide us with the relevant entities. But what account can Semantic Externalism offer of possession conditions for concepts which lack external individuation conditions? It is hard to accept that we should all be hostage to such illusions of content, that in fact we believed nothing when we sincerely assented to 'phlogiston is a gas', and so on.

For such reasons, I imagine, Burge [1982] has tried to envisage a Dry Earth scenario that purports to underpin that it is logically possible to have, say, water thoughts without the existence of water. What he has in mind is not only a case in which an individual has a belief involving the concept water even though there is no water in his local environment of which he holds this belief, but a case in which an individual has a belief involving the concept water even though there exists no water at all in his global environment (cf. fn. 24). What we have is thus a candidate counterexample to Boghossian's Dry Earth. However, care is needed in stipulating the case. What must be required is that there exists sufficient knowledge of chemistry amongst the more informed members of that

individual's community to distinguish the concept water from various twin-concepts. If the community possessed a chemical analysis of water showing it to be composed of two hydrogen atoms for every oxygen atom, both of which they were causally related to, there would be no obstacle to regarding that individual as in possession of the concept water. Likewise, we can stipulate a situation in which the individual could form the concept water, even if there were no other speakers in his environment. What would be required is that he had causally interacted with instances of the natural kind water and furthermore held a minimum of chemically relevant beliefs to distinguish water from other candidate twin-substances. What we cannot have, according to Burge, is a combined situation in which an individual has the concept water even though neither water nor other speakers exist in his environment.

Burge's Dry Earth story is unfortunately contingent upon the assumption that although the concept water is atomic, water itself is a composite natural kind: water is H_2O . It is namely presupposed that the concept water functions in much the same way a composite natural kind concept, e.g. the concept H_2O , functions. It is required that those more skilled fellows of his have had causal encounters with the constituent natural kinds, oxygen and hydrogen, of which water is composed, and then construed a chemical theory about water deploying those atomic, conceptual constituents. Yet this can easily be circumvented: just pick a real atomic natural kind concept, like the concept oxygen, or supposedly any other concept of the natural elements, and then set up a Dry Earth scenario. In that case, the necessary causal encounters with the constituent substances would not be available to the experts to build the chemical theory, and there would be no concept at hand for the individual. What really has to be shown is not that one can have an atomic natural kind concept of a non-existing composite natural kind, but that one can have an atomic natural kind concept of a non-existing atomic natural kind, given what Semantic Externalism says about concept individuation. The former can be made plausible provided the component natural kinds exist and that the individual or his fellow speakers are clever

enough. The latter, however, seems impossible given Semantic Externalism and yet nothing prevents us from running a Dry Earth Argument on such atomic natural kind concepts.²⁶

Let me take stock. I have presented an argument - the Dry Earth Argument - which purports to show that once external individuation is allowed for, object-dependence must also be endorsed. But object-dependence is a form of Semantic Externalism sufficiently strong to sustain the second premise in arguments like (3*)-(5*) of the MC-form. Given that the first premise is a priori warranted and that the second premise is properly interpreted as a conceptual entailment, an a priori warrant for the conclusion can then be deduced. But since we agreed only to accept an a posteriori warrant for the conclusion, we are forced to give up the a priori knowability of either or both of the premises. So, what we effectively have is a *reductio* of Compatibilism, i.e. the view that Privileged Access and Semantic Externalism are compatible doctrines. The problem is that, as we saw in Sec. 1 and 2, both doctrines are independently justified. It would thus be theoretically unsatisfactory if we left the dialectic at this stage. As with any paradox, what is called for is not just negation of one or more of the premises, but diagnosis of why they all seem justified and simultaneously lead to paradox by a seemingly cogent line of reasoning. What must be provided is thus some account of why both doctrines seem so plausible when their conjunction appears to lead to absurdities. I shall be occupied with developing such an account in the following. A close look at the literature indicates that most Semantic Externalists agree that if we have a *reductio* of Compatibilism, then we also have something very close to a *reductio* of Semantic Externalism. They have therefore spent

²⁶ The point is conceded by Burge [1982] and McLaughlin & Tye [1998, pp. 300-2]. I shall have more to say in the following about why Weak Semantic Externalism looks like an unstable position. Note also that Putnam's characterisation of Semantic Externalism from Sec. 1 seems to be in accord with this claim: a wide state is a state that presupposes the existence of objects other than the individual who is in that state. If we suppose that proposition p presupposes proposition q iff p entails q, then other appropriate objects must exist if the individual is in a given wide state.

much energy explaining why Compatibilism holds after all.²⁷ My contention is that once we get clear on the semantic underpinnings, there is a forthright explanation of why Compatibilism has seem so hard to come by. The bad news is that the MC-form has convinced us that Strong Semantic Externalism and Privileged Access are incompatible. But the former is only sustained by the view I called *Referentialism*. So, if Semantic Externalism is detached from Referentialism, then Compatibilism looks less futile. Weak Semantic Externalism incorporates the Twin Earth lesson without renouncing Compatibilism, and the good news is that it is sustained by an appropriately constrained form of what I called *Descriptivism*; or so I shall argue.

²⁷ See for instance Burge [1988], Davidson [1987], Brueckner [1992] and McLaughlin & Tye [1998].

Chapter 2. The Modal Argument

2.1. Semantic and Mental Content

We saw in Chapter 1 that Semantic Internalism and Externalism are theories about how content is individuated. In the first instance, they disagree about how propositional or semantic content is individuated. We can understand the propositional content of a sub-sentential expression as the contribution it yields to determine the propositional content expressed by an assertoric utterance of a sentence¹ in which it occurs, and we can understand the propositional content of a sentence as its *truth-conditions*. If we call the latter the proposition associated or expressed, we can say that the propositional content of an expression is the contribution it yields to determine the proposition associated or expressed by a sentence containing that expression. But since ordinary speakers have and deploy sufficient means to express their propositional attitudes by assent or dissent to sentences, what they thus believe or desire will partially be a function of the propositions associated or expressed by these sentences. What is true about semantic content is true about mental content. It follows, as we saw in Chapter 1, that if Semantic Internalism and Externalism disagree about how semantic content is individuated, they will also disagree about how mental content is individuated. Moreover, we are inclined to think that the mental states themselves are partially individuated by their contents. What makes my belief that *p* the state it is, is partially that it is a belief with the content that *p*. Had my belief that *p* had a different content *q*, I would have had a different belief - the belief that *q* - and so been in a different belief state.

In order to give a thorough evaluation of the arguments for Semantic Internalism and Externalism respectively, one would thus have to consider the question of how the propositional content of an expression is determined on the two views. I said above that the

¹ In Fregean style, we might say the propositional content of complete complex expressions.

propositional content of a sub-sentential expression is its contribution to determining the propositional content - the proposition - of sentences containing it. Now, for non-referring terms that may be all there is to say about its propositional content, but for referring terms, we can give an independent characterisation that pertains to their referents or extensions. The issue about how such terms refer is thus intimately connected with the issue of what constitutes their propositional content. One cannot assess a theory of meaning in isolation from a theory of reference for referring terms.

But before I turn to this task, let me make a notational comment. By 'expression' or 'term' I shall in the following focus exclusively on *singular terms*, e.g. proper names, and *general terms*, e.g. natural kind terms. There are mainly two ways of handling natural kind terms. One is to treat them as singular terms: 'water' refers to the abstract kind H_2O or to the collection of entities which uniquely have the watery properties. The other is to treat them as general terms: 'water' applies to all and only instances of the kind H_2O or to all and only entities which uniquely have the watery properties. General terms are predicative and have extensions, whereas singular terms are referring and have referents. I shall not distinguish between the two ways and nothing I say hangs on which is adopted.²

2.2. Semantic Internalism and Descriptivism

We saw in Sec. 1.1 that Semantic Internalism held that sentences containing natural kind terms have associated with them descriptive truth-conditions: 'water is wet' is true iff the watery stuff is wet. We called such truth-conditions *descriptive*, because the common noun phrase 'the watery stuff' is shorthand for a list of stereotypical, manifest properties - the clear, potable liquid that falls from the sky, flows in rivers, supports life, etc. - that uniquely *describe* paradigmatic samples of water. It follows on this view that the sentence 'water is wet' is associated with the *descriptive proposition* that the watery stuff is wet,

² For more on this, see Salmon [1982, pp. 42-75].

given that we have identified the truth-condition of a sentence with the proposition associated. We also saw that to believe such a descriptive proposition is to have a narrow content belief in the sense we defined that term: despite our environmental differences, my doppelgänger and I both believe that the watery stuff is wet. That is, to believe a descriptive proposition is to have a mental property which supervenes on one's intrinsic properties. On this view, the natural kind term 'water' thus has as its propositional content the descriptive content of 'the watery stuff'. 'Water' is simply a shorthand for 'the watery stuff'.

Why does the Semantic Internalist hold this view? Well, it is essential that *meaning be in the mind* in the way opposed by Semantic Externalism. As a first approximation we can take this dictum to mean that the meaning of an expression is exhausted by what a competent speaker knows when she fully understands that expression.³ In this sense, meaning and understanding are correlative notions such that no ingredient in meaning can outrun the content of the knowledge possessed by a speaker who has full competence with the relevant bits of language.⁴ The Semantic Externalist, however, need not disagree that meaning is *cognitive* in this sense. The disagreement arises only over the kind of knowledge involved. On the Semantic Internalist's account, the kind of knowledge involved is first and foremost explicit, *descriptive De Dicto knowledge*. If queried what the meaning of 'water' is, a competent speaker will on reflection be able to cite sufficiently many watery properties to enable ordinary identification of water, and an ideal speaker will be able to give the entire list. Such *identifying knowledge* will be introspectively available to the speaker. But implicit or practical knowledge may also play a role. To grasp the meaning of an expression could partially consist in having an inarticulate yet manifestable *recognitional capacity*. In some cases, like concepts for secondary qualities - red, sweet or

³ I shall use 'meaning', 'propositional content' and 'semantic content' interchangeably in the following.

⁴ This is familiar from Dummett's writings who seemed to think that this correlation claim was a platitude governing natural language that any viable theory of meaning should respect, see for instance [1978, p. 129; 1991, p. 83].

loud - knowledge of meaning may consist primarily in an ability to recognise a referent if presented with one.⁵ The important point is that the referent of an expression is no part of the meaning of that expression, hence, no *De Re knowledge* of the referent is constitutively involved in understanding that expression. There are two parts to this: (i) knowledge of reference is *not necessary* for knowledge of meaning since it is perfectly possible for someone to fully understand an expression without knowing its referent, indeed without there being one; (ii) knowledge of reference is *not sufficient* for knowledge of meaning since it is not possible for someone to fully understand an expression without knowing more than its reference. Take 'water'. Just as we can imagine a speaker who does not know that 'water' refers to H₂O yet knows enough of the watery properties in order to grasp 'water'-sentences, we can imagine a speaker who does know that 'water' refers to H₂O yet does not know enough of the watery properties in order to grasp 'water'-sentences. Someone ignorant of chemistry may be said to be in possession of the concept water provided she knows enough of the watery properties, and someone ignorant of the watery properties may be said to lack the concept water even if she knows the underlying chemistry. It is no part of the meaning of 'water' that it picks out all and only H₂O on Earth, since on Twin Earth the very same concept picks out XYZ, and on Dry Earth it fails to pick out anything. What referent or extension an expression has is a question about which if any entities in the world happen to satisfy the descriptive conditions that speakers commonly associate with that expression as its meaning. All understanding takes is knowledge of those conditions, but given that they jointly determine reference, understanding will typically issue in an ability to identify a referent were one presented. Competent speakers normally have sufficient - explicit or implicit - means to determine a referent were there to be one. The crucial point is that, on this view, meanings are not essentially hooked up with their referents. Had it turned out that nothing uniquely had the watery properties, we would

⁵ Cf. Dummett [1973, pp. 110-1]. For a caveat, however, see fn. 19.

still have had the concept water. *Descriptive content is thus object-independent.*⁶ The meaning of an expression is therefore that by which the referent or extension is determined; it is a *mode of presentation*, or a *way of thinking*, of the referent or extension. Although they are different substances, water and twin-water are subjectively or qualitatively indistinguishable. Both appear in the same watery manner which is to say that my doppelgänger and I think of twin-water and water respectively in the same way. And, the Semantic Internalist continues, it is how the world appears to us that is relevant when we individuate the contents of our mental states; it is states thus identified which have causal power to explain behaviour. I reach out for a glass in front of me because I desire a glass of water and truly believe there is one in front of me. But had I been confronted with a glass of twin-water, the same narrow content mental states suffice to explain why I also would have reached out for that glass.⁷

But if 'water' simply abbreviates 'the watery stuff', then 'water' must apply to whatever 'the watery stuff' is true of. The watery properties thus play a dual role: they constitute what someone knows when she understands 'water' and they uniquely determine the reference of 'water'. The reference (or extension) of 'water' is determined by its descriptive content to be (the members of) the class of things, of whatever kind, which uniquely have the watery properties. 'Water' picks out all and only watery stuff such that on Earth, where H_2O uniquely has the watery properties, 'water' picks out H_2O , and on Twin Earth, where XYZ uniquely has the watery properties, 'water' picks out XYZ. This referential mechanism can be made clear if 'the watery stuff' is given the well-known quantificational treatment of definite descriptions. 'The watery stuff' picks out a kind of stuff k iff k is watery and nothing else is. So, 'water is F ' is true iff $(\exists x)(Wx \ \& \ (\forall y)(Wy \rightarrow y = x) \ \& \ Fx)$. It is thus possible that different x uniquely has the watery properties and also

⁶ Accordingly, Descriptivism can accommodate what we intuitively wanted to say about empty cases like phlogiston, namely that truth-conditional content can be retained when the world fails to provide us with an extension of our concepts. This point shall prove important.

⁷ There is a lot to say and I shall get back to this issue in Chapter 5.

is F.⁸ Descriptive truth-conditions are not essentially about any particular kind of stuff but whatever happens to be watery.⁹ And what someone knows when she understands 'water is wet' is just those descriptive truth-conditions. We can take the descriptive content of 'water' to be given by some such *meaning-conferring* stipulation:

(S) Let 'water' refer to whatever stuff has watery properties $P_1, P_2, P_3 \dots P_n$.

Now, in the Twin Earth Argument we took Twin Earth to be a remote place in the actual world W_A . So, on this view, 'water' will refer to different substances even within W_A . But we can also take Twin Earth as a counterfactual possible world W_P . Thus understood, 'water' will apply to whatever stuff is watery at the counterfactual Twin Earth. For all W_P , 'water' refers to whatever stuff is watery at W_P . 'Water' is thus a *flexible designator*, i.e. it refers to different kinds of stuff at different W_P .¹⁰ This follows from the quantificational analysis in that a definite description that contains purely general, qualitative expressions is a flexible designator. Its reference at a W_P depends solely upon whatever or whoever uniquely satisfies the descriptive condition at W_P . We might thus say that 'water' is *constant in content*, but *world-relative in reference*. 'Water' means the same - the watery stuff - but has a different extension on Earth and (counterfactual) Twin-Earth. So, on Earth

⁸ Here I take definite descriptions in their attributive use where an object uniquely has to satisfy the descriptions in order to be successfully picked out by them; cf. Donnellan [1966].

⁹ Similarly for other natural kind terms. If 'tiger' is short for a conjunction of descriptive conditions 'the ferocious creatures that have four legs, are black-striped, are cat-like in appearance...', then 'tiger' refers by way of satisfying these conditions to all and only creatures with the 'tigery' properties, i.e. the class of actual tigers.

¹⁰ Alternatively, we could follow Dummett and call 'water' a *quasi-rigid designator*, because, at any W_P , it refers to any epistemic counterpart of the referent of 'water at W_A '. So, 'water' is flexible with respect to which kind of micro-physical stuff it picks out across W_P , but not flexible with respect to which kind of superficial stuff, as it were, it picks out across W_P .

my tokens of 'water' refer to H₂O and on a counterfactual Twin-Earth they refer to XYZ. Water is H₂O on Earth and XYZ on Twin-Earth.

Let us call the view just sketched *Descriptivism*. In general, Descriptivism says that a singular referring term N refers to an object o at a possible world W_P iff o uniquely¹¹ satisfies a set of descriptive conditions ϕ at W_P which are standardly associated by competent speakers with N as its descriptive content. We can think of ϕ as a *cluster* of descriptive properties like the stereotypical ones we listed for water. N is simply an abbreviation of ϕ , and since it can happen that different objects uniquely satisfy ϕ at different W_P, it can also happen that N refers to different objects at different W_P; N is flexible designator.¹² To understand N, as in sentences in which N occurs, is to possess knowledge of ϕ and know how ϕ jointly determine a referent at a W_P. Descriptivism is thus both a *theory of reference*: the reference of N is determined only by virtue of the satisfaction by o of ϕ ; and a *theory of meaning*: the commonly associated ϕ which a speaker must know of if she is to understand N constitutes the meaning of N. In that case, we shall say that N is *purely descriptive*. The reference relation is thus an *indirect* relation between o and N essentially *mediated* by ϕ . We can therefore separate descriptive reference into two parts. First, the relation between N and ϕ is a *relation of conceptual association* performed by competent speakers. (S), remember, was a meaning-conferring stipulation. Secondly, the relation between ϕ and o is a *relation of fit*: o is the referent of N iff o

¹¹ We shall see in Chapter 4 that there are ways of relaxing this uniqueness condition.

¹² One may object with Russell [1994] and friends Neale [1990] that if N is short for a cluster of definite descriptions, and such descriptions are treated quantificationally, then N is not really a referring expression, since a quantified phrase is not a referring expression. One could follow Russell and say that N denotes o in the sense that N describes or is satisfied by o. Proper referring expressions are those that serve merely to designate an object unmediated by descriptive conditions; in Russell's terminology 'logically proper names'. I think, however, that it would be tendentious to reserve the phrase 'referring' to non-Descriptivist designation. The Descriptivist may want to resist the quantificational analysis and instead treat definite descriptions (The F: Fx) as referring expressions in their own right. I shall not give any of 'denote', 'designate' or 'refer' a technical application, but merely use 'refer' throughout to cover all cases of denotation.

uniquely satisfies ϕ . To understand N is exclusively a question about grasping the first relation regardless of whether any o uniquely fits ϕ . To be the referent of N, however, is exclusively a question about o satisfying an independently fixed ϕ .¹³

So far we have assumed that the conditions ϕ are purely descriptive, that is, consist of definite descriptions given in (i) *general, qualitative* terms: the F, the G, the H, etc.¹⁴ But nothing prevents the Descriptivist from embracing other kinds of definite descriptions: (ii) *meta-linguistic descriptions* which are given in the form of 'the bearer of N'; (iii) *secondary descriptions* which are those fallen back on when the primary descriptions fail to pick out a unique object: 'the object I have in mind' or 'the individual believed to be the F'; (iv) *other-dependent descriptions* which defer to more competent speakers' use of N: 'the object which my linguistic community use N to refer to'; (v) *essential descriptions* which are such that they are necessarily satisfied by any object which uniquely satisfies them: 'the object which has biological origin x'; (vi) *rigidified descriptions* which contains indexical devices: 'the actual F' or 'the F at this world'.¹⁵ Take our test case. We have assumed so far

¹³ I shall not embark on any exegesis as to who may fit my characterisation of Descriptivism. It is clear that I have a motley of views originating in Frege's theory of sense [1994a, 1994b], and Russell's theory of descriptions [1994], in mind. A modern protagonist is Dummett. See for instance [1973, pp. 110-1] where he characterises Fregean sense as an associated criterion for recognising a given object as the referent of an expression; the referent, if there is any, being whatever object satisfies that criterion.

¹⁴ I have bypassed the single-description version of Descriptivism: N is simply short for the F. I do not know of anyone who holds this view anymore. It is, to say the least, very hard to hit upon a single descriptive property F such that anyone who understands N knows that if N refers then N refers to the F. More plausible is the cluster-of-descriptions version originating in Wittgenstein [1953, § 79] and Searle [1958]: N is short for a cluster of descriptions, the F, the G, the H, etc. Modern proponents of Descriptivism, e.g. Lewis [1999a, 1999b] and Jackson [1998b], take their starting point in this version.

¹⁵ I shall not deal with (ii)-(v) in detail. Note, however, that if the Descriptivist relied entirely on descriptions of type (ii) and (iii), she would violate Kripke's Non-Circularity Condition [1980, p. 68]. The fact that N is associated with descriptive conditions was supposed to answer the question 'what is it in virtue of which N refers to o?' To say that N refers to o in virtue of o being the bearer of N or that N refers to o in virtue of being the object I believe to be the referent of N, is not to give an informative answer to that question. Note also that although we often, for good reasons, resort to descriptions of type (iv), they cannot, on pain of infinite

that all the watery properties be of the general, qualitative kind - what we have called stereotypical properties. But it is open to the Descriptivist to incorporate descriptions of the other types into the cluster, e.g. the stuff that goes under the name 'water' in my speech community.

Let me give another example of a Descriptive term. According to Descriptivism, an ordinary *proper name* like 'Bill Clinton' is commonly associated with a cluster of descriptions: the current US president, the Democrat ex-senator in Arkansas, the guy who had an affair with Monica Lewinsky, the person known as 'Bill Clinton', the politician married to Hillary Clinton, etc.; in short, the *Clinton-properties*. To understand 'Bill Clinton' is to know sufficiently many of these conditions to be able to identify Bill Clinton. No knowledge of reference is part of knowledge of meaning: one need not know of Bill Clinton that 'Bill Clinton' refers to him in order to understand 'Bill Clinton', and someone could know of Bill Clinton that 'Bill Clinton' refers to him without understanding 'Bill Clinton'. Moreover, these descriptive conditions are at the same time what uniquely determine the reference of 'Bill Clinton', not only at W_A , but at every W_P . Bill Clinton is whoever uniquely has the Clinton-properties, and anyone who uniquely has the Clinton-properties is Bill Clinton. The name 'Bill Clinton' can be taken as a shorthand for the cluster of definite descriptions of the Clinton-properties.

Since, in general, the descriptive content of N is that by which the reference of N is determined at a given W_P , we can represent this content by a function from W_P to extensions:

$$F_1: W_P \rightarrow \text{Extension/Reference}$$

regress, constitute our ultimate means of determining reference. Descriptions of type (vi), of which descriptions of type (v) is a sub-class, will, however, prove very important for the Descriptivist as we shall see in a short while.

To say that N is a *flexible designator* is to say that the values of F_1 differ as it takes different W_P as arguments. The reference of N at a W_P - be it actual or counterfactual - solely depends upon how things are at W_P . Ignoring descriptions of type (vi), 'water' refers to natural kind k at W_P iff k has the watery properties at W_P . 'Bill Clinton' refers to individual i at W_P iff i has the Clinton-properties at W_P . The (traditional) Descriptivist thus endorses *Descriptivist Reference*:

(DR) N refers to o at W_P iff o satisfies ϕ at W_P .

In sum, we have seen that Semantic Internalism and Descriptivism are tightly connected in that the former presupposes some version of the latter. What we shall now see is that there is a correspondingly close connection between Semantic Externalism and Referentialism. Traditionally, Semantic Externalism has been taken to stand to Referentialism as Semantic Internalism stands to Descriptivism. At first blush, this seems right, but I shall eventually argue that the picture is more blurred. It is true that Semantic Externalism follows from Referentialism, but the former does not presuppose the latter: an adequately constrained version of Descriptivism upholds a version of Semantic Externalism.

2.3. Semantic Externalism and Referentialism

We saw in Chapter 1 that the Twin Earth Argument showed that my doppelgänger's tokens of sentences containing 'water' and my tokens of the same type of sentences had different truth-conditions. My doppelgänger's utterance of 'water is wet' is true iff twin-water, i.e. XYZ, is wet, whereas my utterance of 'water is wet' is true iff water, i.e. H_2O , is wet. We dubbed such truth-conditions *singular* due to the occurrence on the right-hand side of the bi-conditional of the very referent of 'water' - as opposed to a set of descriptive conditions

an object has to satisfy in order to be the referent of 'water'. It followed, given our assumptions, that if mental content is individuated by truth-conditions, and mental states by their contents, then the property of having the belief that water is wet is not intrinsic. What can the propositional content of 'water' consist in according to the Semantic Externalist? Well, if an utterance on Earth of a sentence containing 'water' has truth-conditions which are singular with respect to H_2O , then such an utterance must express a *singular proposition*, given that we have identified the truth-condition of a sentence with the proposition it expresses. A singular proposition is a proposition which is *essentially about* a particular object in the sense that the proposition is individuated by that object. Earthly tokens of sentences containing 'water', for instance, are essentially about H_2O . Descriptive content is no part of the content of a singular proposition. The constituents of a singular proposition are the referents of the terms that make up the sentence that expresses that proposition. The singular proposition expressed by an Earthly token of, say, 'water is wet' consists of the stuff water, i.e. H_2O , and the property of being wet: $\langle H_2O; \text{wet-hood} \rangle$.¹⁶ It follows straightforwardly that had H_2O not existed, 'water is wet' could not express a proposition which is singular with respect to H_2O . *Singular content is thus object-dependent*. To see why the Semantic Externalist appears to be committed to singular propositions, suppose an Earthly token of 'water is wet' did express some descriptive property, say, the F. Since we assume that 'the F' expresses a non-singular property that invokes no environment-specific entities, it will be possible for doppelgängers to share beliefs with 'the F' in the embedded clause. So, to the extent that 'water is wet' expresses descriptive properties, one would have to allow for narrow content belief with 'water is wet' as the embedded sentence. It would therefore seem as if a thorough-going Semantic Externalist must endorse the claim that the propositional content of 'water is wet' is nothing

¹⁶ One could adopt a similar notation when assigning singular truth-conditions: 'water is wet' is true iff $\langle H_2O; \text{wet-hood} \rangle$, and thereby emphasise that what occurs on the right-hand side is the stuff - H_2O - and the property - wet-hood - themselves, and not some preferred, or even canonical, mode of presentation of them.

but singular. This means in particular that the natural kind term 'water' cannot be associated with descriptive content, but must refer *directly*, i.e. pick out its referent *unmediated* by any descriptive condition. We shall call such a term *purely Referential*¹⁷ and the view that all singular and general terms are purely Referential for *Referentialism*. A purely Referential general term has as its extension objects of the same kind, or, one might say, it refers directly to the abstract kind. The term 'water', for instance, applies to all and only things that are members of the kind H₂O, and not to the class of watery things as Descriptivism would have it. This means that the contribution a purely Referential term yields to determining the proposition expressed by sentences containing it is its referent: its meaning is its referent.¹⁸ The term 'water' does not abbreviate a conjunction of definite descriptions that express the watery properties; it functions semantically as a mere referring device. If the correlation claim - the meaning of an expression is what someone knows when she understands it - is accepted, it follows that someone who understands a purely Referential term as it occurs in sentences, must have knowledge of its referent that the term refers to it. In order to understand an utterance of, say, the sentence 'water is wet' one has to grasp its singular truth-conditions and to grasp them is to have De Re knowledge of water, i.e. H₂O, that it is wet. What is more, the having of such knowledge seems to require as a necessary condition that one stands in suitable *epistemic relations* to H₂O, since we know that unique identifying knowledge will not suffice. What exactly such knowledge amounts to is unclear and probably varies from concept to concept. However, a minimal constraint must be that

¹⁷ Also known in the literature as directly referential or Millian terms. With Kripke's words [1994, p. 353], a proper name is Millian if "It *simply* refers to its bearer, and has no other linguistic function. In particular, unlike a definite description, a name does not describe its bearer as possessing any special identifying properties."

¹⁸ It is of course possible that a Referential term, as well as a Descriptive term, has other non-semantic ingredients in its meaning: connotations, tone, colour and what not. By 'meaning' and 'propositional content' I understand semantic content, content which is truth-conditionally relevant.

one, or at least a fellow speaker, sustains adequate *causal links* with instances of the kind H₂O, e.g. that one has had perceptual contact with such instances.¹⁹

The same is true of purely Referential singular terms. The proper name 'Bill Clinton' does not refer to Bill Clinton by satisfaction of the Clinton-properties - it refers *directly* to him. 'Bill Clinton' is not short for 'the individual with the Clinton-properties'. This means that the referent of 'Bill Clinton' must enter into propositions expressed by sentences containing the name, i.e. that such sentences express singular propositions. An utterance of 'Bill Clinton is kind' expresses the proposition <Bill Clinton; kindness>, since it has the singular truth-conditions: 'Bill Clinton is kind' is true iff Bill Clinton is kind. It follows that had Bill Clinton not existed, that sentence could not have expressed the proposition it does express. The propositional content of the name 'Bill Clinton' - its meaning - is its referent, since this is its contribution to determining the proposition expressed by sentences containing it. The semantic function of a purely Referential proper name is merely to pick out a referent. Again, if the Clinton-properties do not constitute the meaning of the name, then knowledge of these properties can be no part of what one knows when one understands the name. To understand an utterance of 'Bill Clinton is kind' is to grasp its singular truth-conditions; it is to know of Bill Clinton that he is kind and to possess such De Re knowledge requires, in turn, that one be - more or less - epistemically

¹⁹ The difficulty of spelling out what such De Re knowledge consists in corresponds to the difficulty from Sec. 1.5. of making explicit the external condition E on wide concept possession. The main point is, with Russell [1980], that *knowledge by description*, or propositional knowledge, will not suffice. One must have *knowledge by acquaintance* and it seems that one cannot in this sense *know which* object is being referred to if one has had no epistemic contact with the referent, if, in particular, one lacks *discriminating knowledge*. For discussion see Evans [1982, pp. 89-100], Davies [1981b, p. 97] and Neale [1990, pp. 16-19]. Note how this parallels the fact that singular content is object-dependent and descriptive content is object-independent. One cannot have knowledge by acquaintance if there is nothing of which one is acquainted, but one can have knowledge by description even if nothing happens to satisfy the description(s). The Descriptivist should thus beware of embracing a kind of knowledge or epistemic ability that could not be attained were the world not to co-operate with our linguistic stipulations.

linked up with Bill Clinton. To know the meaning of a name is thus to have knowledge of reference in the sense that one knows of the referent that the name refers to it.²⁰ But if our stipulation from Sec. 2.2.:

(S) Let 'water' refer to whatever stuff has watery properties $P_1, P_2, P_3 \dots P_n$,

is not meaning-conferring, what semantic work does it do? According to Referentialism, all descriptive conditions can accomplish by way of determining meaning, is to fix the reference once and for all of a term being introduced. (S) is a device for introducing 'water' into language - it facilitates the baptising of a kind of substance. The reason that descriptive conditions can play no role in determining reference on subsequent occasions or in counterfactual circumstances is that they confer no propositional content upon the term. (S) is a mere *reference-fixing* stipulation. Once the reference of 'water' has been determined by satisfaction of the watery properties to be H_2O on Earth, then 'water' refers to H_2O on all subsequent occasions and with respect to all counterfactual circumstances. In other words, the natural kind term 'water' rigidly picks out the kind of stuff that has the watery properties at the actual world W_A . On Earth 'water' is a *rigid designator* of the natural kind H_2O , and on Twin Earth, 'water' is a rigid designator of the natural kind XYZ.²¹ As opposed to Descriptivism, we may say that 'water' is *variant in content*, but *world-bound in reference*. On Earth 'water' is set up by my use of (S) to refer rigidly to H_2O . Hence, my tokens of 'water' do not refer to XYZ on Twin Earth even if XYZ equally has all the watery

²⁰ Thus suppose I say: let 'Newman 1' refer to the first child born in the 22nd century. Given that I am not now informationally - or causally - linked up with the referent of 'Newman 1', I am not now in a position to have beliefs that I would express using 'Newman 1'. For disagreement see Kaplan [1989] who thinks that it is possible for a speaker to entertain a singular proposition although its constituents are not epistemically salient to her.

²¹ Donnellan [1966], Kaplan [1989], Putnam [1990, 1996] and Kripke [1980] have all argued that natural kind terms are rigid designators. The claim is almost universally accepted nowadays.

properties, and so 'water' means water on Earth. On Twin Earth, however, 'water' is set up by my doppelgänger's use of (S) to refer rigidly to XYZ. Hence, his tokens of 'water' do not refer to H₂O on Earth even if H₂O equally has all the watery properties, and so 'water' means twin-water on Twin Earth.

In general, Referentialism is thus a denial of the claim that there is one thing - descriptive content - such that it is both what someone knows when she understands a referring term N and is what determines the reference of N with respect to W_P. No single entity plays this dual role in the theory of meaning and the theory of reference. Descriptive content can be dispensed with in a theory of meaning: to understand N as in sentences in which it occurs is to have De Re knowledge of the referent of N and to meet whatever epistemic constraints that involves. The role of descriptive content in a theory of reference is restricted to that of reference-fixing: once the reference of N is fixed at an initial episode of dubbing an object o, there is no further question about what the referent of N is; N directly picks out o at all W_P - unmediated by any descriptive condition ϕ .²² N is *modally rigid* in that the object N picks out at W_A is the object it picks out at all W_P. We can thus define modal rigidity in terms of possible worlds:

(W_P-Rigidity) N is rigid designator iff N refers to the same object o at every W_P at which o exists and not to something else at W_P at which o does not exist.²³

²² There are in fact two separate issues here. One concerns the *descriptive-semantic* question about what semantic value N has in our language; another concerns the *foundational-semantic* question about what makes it the case that N has the semantic value that it has; cf. Stalnaker [1997]. Descriptivism tells us that the semantic value of N is its descriptive content, and answers the second question by saying that descriptive content is what someone knows - implicitly or explicitly - when she understands N. Referentialism, on the other hand, tells us that the semantic value of N is its referent, and answers the second question by a causal theory of reference: my present use of N goes back by some causal chain of communication to an original act of baptising; cf. Kripke [1980] and Evans [1996a].

²³ This leaves it open whether N is a *persistent* rigid designator that refers to o at W_P only if o exists at W_P or an *obstinate* rigid designator that refers to o at W_P even if o does not exist at W_P. I shall not take a stand on

Thus if we represent singular content by our function F_1 from W_P to extensions/referents, then the values of F_1 will remain constant as it takes different W_P as arguments. To say that N is a rigid designator is to say that the reference of N at a W_P - be it actual or counterfactual - solely depends upon how things are at W_A - bracket questions about existence at W_P (cf. fn. 22). The Referentialist endorses *Referentialist Reference*:

(RR) N refers to o at W_P iff o satisfies ϕ at W_A .

With our examples, 'water' refers to natural kind k at W_P iff k has the watery properties at W_A . Likewise, if we suppose 'Bill Clinton' to have been introduced by analogy to (S), 'Bill Clinton' refers to individual i at W_P iff i has the Clinton-properties at W_A .²⁴ Note also that on both the Descriptivist and the Referentialist accounts, reference at W_A is uniquely determined by how things are at W_A . They will thus agree on the reference-conditions for N with respect to W_A :

(W_A -R) N refers to o at W_A iff o satisfy ϕ at W_A .

this issue. Kripke [1980, pp. 21, 78] assumed that ordinary proper names were of the latter kind: if I say 'suppose Hitler had never been born', I refer to Hitler at a W_P at which he does not exist. See also Kaplan [1989] and Salmon [1982, pp. 32-40]

²⁴ The Referentialist need not claim that all names necessarily are rigid designators. As Kripke [1980, pp. 79, 91, 94] says, we could take 'Jack the Ripper' to refer flexibly to whoever was the murderer of several prostitutes in 1890's London, such that if Jones was the actual murderer, he would be Jack the Ripper, but had James been the murderer, he would have been Jack the Ripper, etc. I shall henceforth use 'Jack the Ripper' in this manner. If the reader disagrees, assume the name has such a use for the sake of argument. Kripke's contention is merely that Descriptive names *actually* form a minor sub-class of ordinary proper names the majority of which are purely Referential.

'Water' refers to natural kind k at W_A iff k has the watery properties at W_A , and 'Clinton' refers to individual i at W_A iff i has the Clinton-properties at W_A . (W_A -R) highlights the fact that the reference of N , taken as purely Descriptive, and the reference of N , taken as purely Referential, coincide when it comes to W_A .²⁵ The class of watery things at W_A is identical to the instances at W_A of the natural kind H_2O , and the individual who has all the Clinton-properties at W_A is the individual who bears the purely Referential name 'Bill Clinton'. This point will prove important: the difference between Descriptivism and Referentialism with respect to reference only arises when it comes to counterfactual worlds.

Let me finally emphasise how Referentialism underpins, not only the assumption made in the Twin Earth Argument that Earthly and Twin Earthly tokens of 'water'-sentences have different singular truth-conditions, but also *object-dependence*: were I on Dry Earth, my tokens of such sentences would lack singular truth-conditions. Under these imagined circumstances there is no answer as to *which kind of stuff* - which kind of micro-physical conditions - would make my tokens true or false, and so no answer as to *which proposition* I could thereby have expressed. This strongly suggests that our Dry Earth Argument is sound, or more precisely, that if Semantic Externalism is backed up by Referentialism, then there is no stable intermediary position between Semantic Internalism and Strong Semantic Externalism.

²⁵ It may be objected that Referentialism is neither committed to (RR) nor to (W_A -R) given that there are versions of causal theories of reference which account for reference-fixing without invoking descriptive conditions. Maybe so. The point is that, following Kripke [1980] and others, it is clearly a possibility that a term may have its reference fixed by satisfaction of some descriptive condition without that condition entering into the propositional content of the term.

2.4. The Modal Argument

In the previous two Sections we have seen the close connections between Semantic Internalism and Descriptivism, and Semantic Externalism and Referentialism. Which doctrine about mental content is adopted will depend on which doctrine about semantic content is viable. This is not much news to anyone. Putnam [1990, 1996], for one, was explicit that his attack on Semantic Internalism was paralleled by the range of arguments that Kripke [1980] almost simultaneously had launched against Descriptivism. To assess whether to take the conclusion of the Twin Earth Argument into account, one should evaluate the arguments that support the presuppositions of that Argument. To this task I shall turn in a moment. But before that, consider what kind of arguments may be brought to bear on either Descriptivism or Referentialism. We are interested in arguments that purport to show what the propositional content of a referring term is. One way to do that is to have an argument that says something about the proposition expressed by an atomic sentence - a sentence free from intentional and modal vocabulary - containing that expression. Then we could use:

(Compositionality of Content) The propositional content of a sentence is uniquely determined by the content of its constituent expressions and their syntactic manner of combination,

to deduce a conclusion about the content of the contained expression. Again, by proposition I shall understand what is expressed by an utterance of an atomic sentence, and what is thus expressed is both what someone knows when she understands the sentence and also the conditions under which it is true. In the following I will be concerned with the question of which properties of what is expressed by an utterance of a sentence are relevant for determining the propositional content of that sentence. It is striking that the same line of argument is adopted by both doctrines: the *Content - Property (CP)-Argument*. The CP-

Argument is a *reductio* and runs like this: suppose that two singular terms α and β coincide in content. Then, by (*Compositionality of Content*), two atomic sentences

(1) $S(\alpha)$

(2) $S(\beta)$,

whose only difference is that (1) contains α and (2) contains β , would also coincide in content. Consequently, if the content expressed by (1) were identical to that expressed by (2), they would both, by Leibniz's Law, have a certain property ψ . But, *ex hypothesi*, the content expressed by (1) has ψ , i.e.

(3) ' $S(\alpha)$ has ψ '

is true, while the content expressed by (2) does not, i.e.

(4) ' $S(\beta)$ has ψ '

is false, and so the content expressed by (3) is not identical to that expressed by (4). Hence, by (*Compositionality of Content*), the contents of (1) and (2) must differ, hence, the contents of α and β must differ.

Now, the dispute between Descriptivism and Referentialism concerns which instances of the CP-Argument are persuasive, i.e. which ψ are the *relevant* content-individuating properties. To begin with, the *truth-value* of an atomic sentence, or for its sub-sentential expressions, extension or reference, is a *minimal* ψ that any satisfactory account of content must respect. Given the identification of content with truth-conditions for such sentences, (1) and (2) must express different contents if they receive different truth-values, and likewise for (3) and (4) provided ψ is a relevant content-sensitive property. Or if two terms refer to different objects in the same context, then they must differ in content. It is equally obvious that there are ψ which can be excluded out of hand such as 'is expressed by linguistic expression π ', 'gives rise to conversational implicature υ ', or 'is informative to an incompetent speaker'. Let me illustrate. Suppose it is claimed that:

(5) Vixens are female foxes,

and

(6) Vixens are vixens,

which intuitively have the same truth-conditions, coincide in content. Surely we would want to rule out instances of the CP-Argument to the effect that (5) and (6) express different contents for the reason that the two contents are expressed by different sentences, or that some conceptually ignorant speaker might take different cognitive attitudes towards (5) and (6). If such ψ were relevant content-individuating properties, no two terms would ever have the same content. Between platitudinous ψ 's and irrelevant ψ 's, there is room, however, for argument that some instances of the CP-Argument are plausible.²⁶ In particular, Referentialism claims that *modal semantic value* is one such content-sensitive property. First some background.

We have seen that according to Descriptivism, natural kind terms like 'water' or 'tiger' are shorthand for a list of definite descriptions which we abbreviated as 'the watery stuff' and 'the tiger-properties'. It follows that if that list contains descriptions of type (i) - (iv) only, the terms 'water' and 'tiger' will be modally flexible, i.e. they will apply to different kinds of objects at different W_p . Take the Twin Earth Argument from Chapter 1, but now think of Twin Earth as a counterfactual W_p . The claim would then be that Descriptivism gets the truth-conditions wrong with respect to such a counterfactual Twin Earth. According to Descriptivism, Earthly utterances of 'water is wet' are true at W_p iff the watery stuff at W_p is wet, so in particular, they are true at Twin Earth iff the watery stuff on Twin Earth is wet.²⁷ But these are not the conditions under which Earthly tokens of 'water'-sentences are true; only the obtaining of H_2O -facts can make such tokens true. The intuition is thus that 'water' is a rigid designator of H_2O , for had 'water' referred to XYZ at

²⁶ It is also striking that most if not all CP-Arguments are *negative* in their intent. It is well-known that it is very hard to give necessary and sufficient conditions for two expressions to coincide in content. Co-extensionality, for instance, is clearly no good as 'creatures with heart' and 'creatures with kidney' illustrate, and, as we shall see in the next Chapter, even necessary co-extensionality may not suffice.

²⁷ Again, by 'water is wet' is *true at W_p* , I mean that the proposition actually expressed by 'water is wet' is true at W_p , and not that 'water is wet' would express a true proposition were it uttered in W_p .

some W_P , one would have to consider XYZ-facts when evaluating the truth-values of Earthly utterances of 'water'-sentences with respect to those W_P . What the modal rigidity of an expression shows seems to be that associated descriptive properties constitute neither a *sufficient* nor a *necessary* condition for determining the reference at W_P . The former is illustrated by our counterfactual Twin Earth: XYZ has all the watery properties and yet is intuitively not water. Or imagine a counterfactual W_P where there are no tigers, but a qualitatively indistinguishable species of felines who nevertheless are genetically different from tigers. Such twin-tigers are not tigers. The latter is also conceivable. Imagine a counterfactual Twin Earth on which H_2O lacks the watery properties. Given that H_2O uniquely has the watery properties on Earth, intuitions have it that even H_2O on this Twin Earth is water. Similarly, we can imagine a W_P at which something is a tiger while not having the tiger-properties. In both cases, I take it, our intuitions are informed by the kind of micro-physical reductions I mentioned in Sec. 1.2: once physical science discovers that H_2O occupies the causal role played by water, it discovers the necessary identity that water is H_2O .²⁸

In fact, as Putnam [1970] pointed out, it may well happen that the stereotypical properties are defeated at W_A . We know of, or could easily imagine without change in the laws of physics, abnormal members of natural kinds, e.g. green lemons that never turn yellow, or non-striped, three-legged tigers, etc. There is, however, an important difference between what can happen at W_A and at counterfactual W_P . At W_A it is - next to - physically impossible that something can be a member of a natural kind and yet lack *all* typical descriptive properties. The Descriptivist could therefore insist that for something to

²⁸ Putnam [1996] did not rest content with possibilities grounded in such theoretical identifications. He thought that pencils could have turned out to be organisms, cats could have been robots, paediatricians might have been a Martian species. It is not clear that we have firm intuitions about how to describe such possible cases. For scepticism see Dummett [1973, pp. 143-6]. In any case, note that also the Descriptivist has the conceptual resources to accommodate such possibilities: 'being human' need not be a cluster-property associated with 'paediatrician'. Cf. also Schwartz [1996].

be a member of a natural kind, it need only have *sufficiently many* of these properties in order to count as a member of that kind. This move would get around problems with uniqueness at W_A - provided we knew how many properties were enough. But the Referentialist can push her case: at counterfactual W_P we can imagine instances of natural kinds that lack every single standard characteristic, e.g. an H_2O -sample which has none of the watery properties is nevertheless water. Due to the modal rigidity of natural kind terms, they pick out at every W_P whatever stuff is substance-identical with the stuff that has the associated descriptive properties at W_A , whether or not it, or anything else, has those properties at those W_P . So, since there are W_P at which something is, say, water without having the watery properties, and W_P at which something has the watery properties without being water, these properties do arguably not enter into the propositional content of 'water'. It can be no part of the propositional content of such terms to have associated any such properties; their referents, if they have any, seem to provide all their content.²⁹

The same is true of proper names. Suppose 'Aristotle' is a flexible designator, because simply short for 'the teacher of Alexander', etc. Then suppose W_1 is a world just like W_A except Plato taught Alexander, etc., and suppose W_2 is a world just like W_A except nobody taught Alexander, etc. According to Descriptivism, 'Aristotle' would then refer to Plato at W_1 , and not to Aristotle at W_2 . But that has highly counter-intuitive consequences. Suppose I utter 'Aristotle was born in Stagira'. Then what I have said would be true at W_1 because of the fact that Plato was born in Stagira at W_1 , and what I have said would lack a truth-value at W_2 because of the fact that nobody taught Alexander, etc. at W_2 . Surely, what I have said is false both at W_1 and W_2 . So, 'Aristotle' cannot be a shorthand for 'the teacher of Alexander', etc. The fact that 'Aristotle' rigidly picks out its

²⁹ The Referentialist need not render descriptive content altogether redundant in a theory of meaning for natural kind terms. Putnam [1970, 1978, 1990, 1996] and McGinn [1999] have proposed hybrid theories which incorporate such stereotypical, but not semantically relevant, information; see Chapter 4 for more.

referent seems to imply that it can have as no part of its propositional content any associated descriptive properties; its referent, if it has any, provides all its content.

The last two paragraphs contain what has come to be known as the *Modal Argument*.³⁰ It is an instance of the CP-Argument which deploys modal properties as content-individuating: if N were short for the F, the G, the H, etc., in short ϕ , then N and ϕ would have the same modal properties. In particular, 'N might not have been ϕ ' would be false since equivalent to ' ϕ might not have been ϕ ' which is false. But 'N might not have been ϕ ' is true due to the different modal properties that N and ϕ have - N is modally rigid whereas ϕ is modally flexible. So, N and ϕ must differ in content. In other words, were N simply an abbreviation of ϕ , one should be allowed to substitute the two expressions in modal contexts *salve veritate*. But widespread intuition has it that such substitution fails to preserve truth-value when those expressions are actually, but not counterfactually co-referring, and so N does not abbreviate ϕ . More formally, the CP-Argument runs as follows: suppose the propositional content of N is given by the definite descriptions ϕ - N is just a convenient shorthand for ϕ . Then, by (*Compositionality of Content*), the two atomic sentences

(7) N is N,

(8) N is ϕ ,

should also coincide in content. If so, then, by Leibniz Law $[\alpha = \beta \ \& \ E(\alpha)] \rightarrow E(\beta)$, the contents expressed by (7) and (8) must have the same content-individuating properties. In particular, they should have the same *modal properties* where this is understood to mean that the sentences that express those contents must embed similarly inside the scope of a modal operator. Hence,

(9) N might not have been N,

and

³⁰ Originally due to Kripke [1980]. For a forceful recent statement of the Modal Argument see Soames [1998].

(10) N might not have been ϕ ,

should also coincide in content. Yet we are strongly inclined to say that (9) is false while (10) is true, since there are W_P at which N is not ϕ , but no W_P at which N is not N. We know that there will be W_P at which N and ϕ are not co-referential, because we know that N is modally rigid and ϕ modally flexible. It follows that a difference in truth-value entails a difference in content, and so that (9) and (10) must differ in content, hence, so must (7) and (8), and therefore also N and ϕ . The Modal Argument applies to all referring terms: substitute N and ϕ for 'Aristotle' and 'the teacher of Alexander', etc., 'water' and 'the watery stuff', 'tiger' and 'the tiger-properties', and so on.

The Modal Argument seems - on the face of it - persuasive. What could the Descriptivist say in response? The first worry that comes to mind concerns an *operator-shift fallacy*. It is clear that the validity of the Argument depends on whether N takes the same scope in (9) and (10) with respect to the modal operator 'might not'. Suppose first that, as we normally do, N is read as taking *wide scope* with respect to 'might not'. Thus (10) is true on that reading: $\exists x(x = N \ \& \ \Diamond \neg \phi N)$. Is (9) true on that reading: $\exists x[x = N \ \& \ \Diamond \neg (N = N)]$? It seems not, but the Descriptivist will insist that (9) is semantically equivalent to:

(11) ϕ might not have been ϕ ,

which is true if read with wide scope: $\exists x[\forall y (\phi y \leftrightarrow x = y) \ \& \ \Diamond \neg \phi x]$. There are W_P at which ϕ at W_A is not ϕ at W_P . Thus read (9) is in fact true. Compare with 'the survivor might have died' which is true if read (of the survivor, possibly, he died), but false if read (possibly, the survivor died). So, there is no way (9) can be false *and* (10) true on a wide scope reading. So, perhaps N in (10) should have *narrow scope* with respect to 'might not': $\Diamond \exists x(x = N \ \& \ \neg \phi N)$. Thus read, (10) seems true. Is (9) then false on a narrow scope reading: $\Diamond \exists x[x = N \ \& \ \neg (N = N)]$? It clearly is. Even the Descriptivist who maintains that (9) is equivalent to (11) must accept that (11) is false if the description-operator ϕ takes narrow scope: $\Diamond \exists x[\forall y (\phi y \leftrightarrow x = y) \ \& \ \neg \phi x]$. There are no W_P at which ϕ at W_P is not ϕ at W_P .

But the Descriptivist will object that so is (10). On a narrow scope reading, (10) says that there is a W_P at which N is not ϕ . But given that at W_P N picks out whatever ϕ picks out, there can be no W_P at which N is not ϕ . It follows, or so runs the thought, that there is no way (9) and (10) can differ in truth-value on a narrow scope reading either. So, unless the Modal Argument is to trade on an illicit shift of scope, there is no way (9) and (10) can receive different truth-values, and so no reason has been provided why they should not have identical contents.

I think this response is consistent on the part of the Descriptivist. To presuppose that (9) has no true reading is to presuppose that N picks out the same referent inside as well as outside the scope of the modal operator, i.e. that N is a rigid designator. To make this presupposition against someone who thinks that N is modally flexible is clearly question-begging. So, we are trapped in a deadlock. The Referentialist uses the modal rigidity of N to show that N and ϕ must differ in content. The Descriptivist denies that N is modally rigid and insists on its descriptive content ϕ . What must be proved, and not just presupposed, is therefore that N is a rigid designator. To provide a knock-down argument is, however, notoriously difficult. Nevertheless, I think the intuitions in favour of rigidity are overwhelming. It is just highly counter-intuitive to say that (9) has a reading on which it is true and that (10) has a reading on which it is false. How can an utterance of 'Aristotle might not have been Aristotle' be understood as true? Even if put this way: (of Aristotle, possibly, he is not Aristotle), (9) is still false. There is no W_P at which Aristotle at W_A is not Aristotle at W_P , and the reason we are inclined to say this is our feeling that 'Aristotle' is modally rigid. And how can an utterance of 'Aristotle might not have been the teacher of Alexander, etc.' be understood as false? Even if put this way: (possibly, Aristotle is not the teacher of Alexander, etc.), (10) still seems true. There is a W_P at which all things true of Aristotle at W_A are false of Aristotle at W_P . Again, it flies against intuition to say that we do not speak about the same man. It is granted that if Descriptivism is true, then they can be understood in these ways, but so much the more reason to think it is not. In other words,

were Descriptivism true, (9) would be ambiguous between a true wide scope reading and a false narrow scope reading in the way (11) is, but (9) is - we all agree - unambiguously false.³¹

The foregoing squares with what Kripke said about rigidity. He never presented an explicit and cogent argument to the effect that referring terms were modally rigid, but appealed to speakers' intuitions about the behaviour of such expressions in modal contexts. To decide whether any given term is modally flexible or rigid, Kripke [1980, pp. 48-9] proposed an intuitive test - the *Gödel-Schmidt Test* - which, slightly modified, says:

If N is a referring term and ϕ is an associated descriptive condition, then if 'N might not have been ϕ ' is false, when N takes narrow scope with respect to the modal operator, then N is modally flexible.

Thus 'Jack the Ripper might not have been the murderer of several prostitutes in 1890's London' is false, when 'Jack the Ripper' takes narrow scope with respect to 'might not', i.e. $\Diamond \exists x(x = j \ \& \ \sim jM)$ is false, and so is modally flexible. The scope-constraint on N is important for without it 'Jack the Ripper' would fail the test: $\exists x(x = j \ \& \ \Diamond \sim jM)$ is true. Not so for 'Gödel'. 'Gödel might not have been the discoverer of the incompleteness of arithmetic' is true when 'Gödel' takes narrow scope, $\Diamond \exists x(x = g \ \& \ \sim gD)$; indeed one may take the fact that the wide scope reading, $\exists x(x = g \ \& \ \Diamond \sim gD)$, is also true as evidence that

³¹ Our intuition is, in other words, that referring expressions are scopeless, i.e. do not exhibit a De Re - De Dicto distinction in modal contexts; cf. Peacocke [1975]. There is a well-known Descriptivist response in the literature due to Dummett [1973]. The idea is to reduce our intuitions about rigidity to syntactic scope-conventions: referring expressions are shorthand for a cluster of wide-scope definite descriptions. As I have tried to show, this view has counter-intuitive consequences, but I shall not argue here that it is simply a non-starter. Let alone the problems posed by non-modal sentences where there is no room for scope-maneuvres. We have the intuition that (9*) 'If N exists, then N is not N' is necessarily false, but that (10*) 'If N exists, then N is not ϕ ' is possibly true, but how can such basic sentences differ in W_P -truth-value on this view when there is no modal vocabulary about which to make scope-distinctions? This *Strengthened Modal Argument* was advanced by Kripke [1980, pp. 10-2], as a reply to Dummett's lengthy and very subtle discussion, and has been developed by Stanley [1999a] and Soames [1998].

'Gödel' always semantically takes wide scope. Accordingly 'Gödel' satisfies (W_P -Rigidity): 'Gödel' refers to Gödel at W_P where Gödel exists, and does not refer to anyone else at W_P where Gödel does not exist. At a W_P where, say, Schmidt uniquely invented the incompleteness of arithmetic', Gödel, and not Schmidt, will be the bearer of 'Gödel', and at a W_P where Gödel failed to discover anything, 'Gödel' still refers to Gödel.³²

Do natural kind terms pass the test? No! An utterance of 'water might not have been the watery stuff' is not false when 'water' takes narrow scope with respect to 'might not', and so 'water' is a rigid designator. Formally: $\Diamond \exists x(x = w \ \& \ \sim Wx)$, where 'W' stands for 'being watery' and 'w' for 'water', is true. To say that there is a W_P at which water - the watery stuff at W_A - is not watery is to say that something might be H_2O and yet lack the watery properties. To repeat, this should carry little conviction as an incontestable proof of the rigidity of 'water'. Someone - with admittedly rather deviant intuitions - may deny the truth of 'water might not have been the watery stuff' when read narrowly on the grounds that if 'water' is short for 'the watery stuff' then our rendition is semantically equivalent to $\Diamond \exists x[\forall y (Wy \leftrightarrow x = y) \ \& \ \sim Wx]$ which is false.

Why do we think referring terms are modally rigid? What grounds our intuitions that, say, 'Bill Clinton' is a rigid designator? Well, suppose I say:

(12) Had Bill Clinton died at birth, Bill Clinton would not have slept with Ms. Lewinsky.

If 'Bill Clinton' had been modally flexible with respect to the Clinton-properties, then it would be unclear what the reference of the second occurrence of 'Bill Clinton' was. But the fact that it is not shows that we use 'Bill Clinton' to talk about Bill Clinton even under circumstances in which he failed to do what he actually did. Or consider an utterance of:

³² As I said in fn. 24 I have assumed that 'Jack the Ripper' is modally flexible, but that 'Gödel' is modally rigid. I am sympathetic to someone who disputes the former; the example is only supposed to illustrate how the Gödel-Schmidt Test could be passed. To be sure, Kripke [1980] never claimed that our referring terms must be purely Referential or anything like that; 'Gödel', for instance, is actually a rigid designator, but we might have stipulated otherwise.

(13) Had Plato done all the famous things Aristotle did, Aristotle would not have taught Alexander.

It is clear that the second occurrence of 'Aristotle' does not refer to Plato despite the fact that Plato - under these counterfactual circumstances - accomplished many of the things Aristotle did at W_A . The fact that we deploy proper names to talk about their actual referents with respect to counterfactual circumstances strongly suggests that the purpose of rigid designation in natural language is to facilitate *counterfactual reasoning*. Such reasoning is essential *inter alia* to the practice speakers engage in when they rationalise about their own behaviour and evaluate that of others. Take our examples. The reason Bill Clinton had an affair with Ms. Lewinsky was that, say, given his oppressed childhood, he could take advantage of his position in the White House when his wife paid him insufficient attention. Had Bill Clinton had a different upbringing, wife and job, he would not have had the affair. And the reason Aristotle taught Alexander was, say, his position at the Academy and his fame as a philosopher. Had someone else had Aristotle's job and written Aristotle's books, then he would have been Alexander's teacher. Without rigidity we could not be sure we were talking about the *same* individual under counterfactual hypotheses about what might happen to an individual we had descriptively identified at W_A . If those conjectures involved substantial changes in what is true of the individual at W_A , it would be a genuine possibility that either reference failed or changed under such circumstances, and the explanatory power would be lost. The same is true of natural kind terms. Consider an utterance of:

(14) Had water not fallen from the sky and flowed in lakes and rivers, water would not have had a global impact on vegetation.

The truth of this subjunctive conditional depends on the two occurrences of 'water' being co-referential. Modal rigidity enables us to fix on the *same kind of stuff* under various hypotheses about what would have happened had water not had the stereotypical properties it has at W_A . That way, we can provide causal explanations: water has a global impact on

vegetation because it falls from the sky and flows in lakes and rivers. Without rigidity we could not envisage how the world would have looked had water not had those watery properties, since we would not know exactly which kind of stuff we were talking about under those counterfactual circumstances.

As we noted earlier, the rigidity of 'water' presupposes that some sort of *sameness-relation* obtains. Some fact must make it the case that we are talking about the same stuff under these hypotheses, i.e. that something is called 'water' at W_P only in virtue of being *substance-identical* to what we call 'water' at W_A . But what are the criteria for sustaining this same-substance relation across W_P ? Given what we have said about counterfactual cases, what constitutes the criteria must be the theoretical identifications of manifest natural kinds with underlying micro-physical kinds. If water is H_2O at W_A , then it is not possible that water is not H_2O . What is possible is that some epistemic counterpart of water - some watery stuff - turns out not to be H_2O , but that is not to say that water, if H_2O , might turn out not to be H_2O .³³ It is the micro-physical identification of water which allows for the possibilities that something non-watery could be water and something watery not water. So, if 'water' is rigid with respect to its micro-structure, then one can safely make counterfactual changes in various manifest properties without losing the identity of the stuff reasoned about since it is simply fixed by its micro-structure. Micro-structure is important, as I said in Sec. 1.2., because of the scientific role of a natural kind concept. We think that manifest natural kinds like water, tiger or gold have their micro-physical structure with necessity, because we think it is their underlying structure which

³³ Note an interesting implication of this: if the Descriptivist allows for descriptions of type (v) to enter into the cluster, i.e. if she writes in *essential* properties in conjunction with the superficial ones, then a Descriptive natural kind term will satisfy the condition for modal rigid behaviour. We can imagine such *modally neutral* terms, Dummett [1981, p. 585], set up to refer to whatever stuff has micro-physical property P regardless of whether the stuff exhibits descriptive symptoms. The drawback, however, is a commitment to the view that knowledge of micro-structure is part of what someone knows internally when she understands natural kind terms, and that seems hard to accept. For more on the whole issue about whether any substantial form of essentialism flows from Referentialism, see Salmon [1982].

causally explains why they have the particular macro-physical properties they have: why they obey certain physical laws and why certain physical law-like behaviour can be predicted, etc.³⁴ If natural kind terms are rigid with respect to the micro-structures of the kinds they pick out, they can serve such purposes of our scientific practice.³⁵

2.5. Rigidification

Let me sum up. We have seen that no viable account of referring terms *N* can render them modally flexible. To deny the modal rigidity of *N* squares badly with speakers' intuitions about their behaviour in modal contexts and such intuitions are informed by the purposes of our linguistic practices involving *N*. But it seems to follow from the Modal Argument that if *N* is modally rigid, then *N* cannot have descriptive content ϕ given by a cluster of definite descriptions, since there would then be W_P at which *N* and ϕ were not co-referential, hence not inter-substitutable *salva veritate* in all modal contexts. So, given that *N* and ϕ have

³⁴ In [1996] Putnam took micro-structure as the criterion for substance-identity, but later [1990] he included physical law-like behaviour. Presumably, physical behaviour supervenes on micro-structure such that if two substances obey different physical laws, e.g. have different boiling points at sea level, then the two substances must have different micro-structures; at least if they are in nomologically identical W_P . This is compatible with two substances obeying the same laws while having different micro-structures. Putnam [1990, p. 69], however, accepts the stronger claim that "...differences in micro-structure invariably (at the actual world) result in differences in lawful behaviour." Depending on what is meant by 'lawful behaviour', this claim, unlike supervenience, could threaten the coherency of the Twin Earth story: if the physical behaviour of H_2O and XYZ is sufficiently different, then twin-water is no longer an epistemic counterpart of water, and so my doppelgänger and I would no longer hold the same narrow content beliefs with respect to the watery properties. In Putnam's story, remember, Twin Earth was supposed to be a remote planet at W_A , and so the laws of nature should be the same as on Earth, but H_2O and XYZ were also supposed to be different substances. He did think [1990, p. 70], however, that if Twin Earth was taken as a counterfactual W_P which obeyed different laws of nature, then the criterion for substance-identity would just be micro-structure.

³⁵ Although the Modal Argument is conceived as the most forceful Referentialist argument against Descriptivism, it is often advanced in conjunction with other CP-Arguments. I shall briefly discuss the Semantical Argument in Sec. 4.4 and the Epistemological Argument in Sec. 4.5.

different modal properties, they must also differ in content. What the Modal Argument thus purports to show is that if two expressions differ in modal properties, then they must *eo ipso* differ in content, i.e. that modal semantic value is a relevant content-individuating property. So, it seems to follow from the Modal Argument that *if N is modally rigid, then N is non-Descriptive*, i.e. purely Referential. It goes without saying that this result is good news for Referentialism. As we saw in Sec. 2.3., this view endorses the claim that an utterance of a sentence containing a non-empty rigid designator has singular truth-conditions. Referential terms serve merely to refer, they are rigid designators devoid of descriptive content.

The Descriptivist should at first blush be reluctant to accept any straightforward entailment from modal rigidity to lack of descriptive content. The fact that language contains *de facto* rigid definite descriptions provides counter-examples to this entailment. Consider: 'the smallest even prime' which happens to be true of one and the same object at every W_P . Mathematical and logical descriptions do all satisfy:

(W_P -Rigidity) N is rigid designator iff N refers to the same object o at every W_P at which o exists and not to something else at W_P at which o does not exist,

but they are clearly purely Descriptive expressions.³⁶ What the Referentialist has in mind by a 'rigid designator' is not really (W_P -Rigidity), but what Kripke [1980, p. 21] calls *de jure* rigidity, "...where the reference of a designator is *stipulated* to be a single object, whether we are speaking of the actual world or of a counterfactual situation...". Take an utterance of the sentence:

(15) Aristotle was fond of dogs.

The Referentialist contention is, with Kripke's words [1980, p. 6], that:

³⁶ Since I take it that if mathematical objects exist, then they necessarily exist, a description like 'the smallest even prime' will refer *de facto* to the number 2 at all W_P *tout court*.

"...there is a certain man - the philosopher we call 'Aristotle' - such that, as a matter of fact, [15] is true iff *he* was fond of dogs. The thesis of rigid designation is simply... that the same paradigm applies to the truth conditions of [15] as it describes *counterfactual* situations. That is, [15] truly describes a counterfactual situation if and only if the same aforementioned man would have been fond of dogs, had that situation obtained."

Or consider a very recent, concise statement of the same line of thought, Soames [1998, p. 2]³⁷:

"...our ultimate ground for thinking that the name *Aristotle* is a rigid designator is our conviction that there is a certain individual *x*, such that for every possible world *w*, the proposition that Aristotle was [fond of dogs] is true at *w* iff *x* was [fond of dogs] at *w*..."

What is in play here is thus a notion of rigidity defined not in terms of W_P as (W_P -Rigidity), but in terms of singular content:

(Singular-Rigidity) *N* is a (non-empty) rigid designator iff there is an object *o* such that for every W_P the proposition expressed by '*N* is *F*' is true at W_P iff *o* is *F* at W_P .

It follows trivially that if *N* is modally rigid by (Singular-Rigidity), then *N* is purely Referential, since the propositional content of *N* will be provided by its referent if it has any. It also follows that if *N* is purely Referential, then *N* contains no descriptive information that could be true of different objects at different W_P ; *N* would serve merely to refer to the same object at all W_P . So, if *N* is rigid by (Singular-Rigidity), then *N* is also rigid by (W_P -Rigidity). It is not true, however, that if *N* is rigid by (W_P -Rigidity), then *N* is also rigid by (Singular-Rigidity).³⁸ De Facto rigid descriptions are counter-examples to

³⁷ See also Peacocke [1975], Davies & Humberstone [1981a] and Neale [1990, Chapter 2].

³⁸ See also [1981, pp. 559-60] where Dummett distinguishes between two Kripkean thesis: proper names are rigid designators and proper names lack descriptive content.

that entailment. So, (Singular-Rigidity) is a *stronger* notion of modal rigidity than (W_P -Rigidity), and a notion that the Descriptivist better not endorse. The problem is, however, that (Singular-Rigidity) seems to be well motivated. The truth-condition of an utterance of (15) does seem to be singular: (15) is true at W_P iff Aristotle was fond of dogs at W_P . It certainly does not have these descriptive truth-conditions: (15) is true at W_P iff the teacher of Alexander *at W_P* was fond of dogs at W_P . To see why not, suppose Plato taught Alexander but disliked dogs at W_1 . Then (15) is false at W_1 because of Plato's distaste for dogs at W_1 . Although we have the intuition that (15) is false at some W_P , this is never due to anyone but Aristotle's opinions about dogs at W_P . What anyone else thought of dogs is utterly irrelevant when evaluating the truth-value of (15) at W_P ; only Aristotle - the man himself - enters the picture. Or take 'water'. The reason we think 'water' is a rigid designator is that - modal - sentences containing 'water' always depend upon how things are with H_2O . That is, Earthly utterances of 'water'-sentences express truth-conditions which are singular with respect to H_2O -stuff, i.e. they are essentially about H_2O -stuff: 'water is wet' is true at W_P iff H_2O is wet at W_P . In contrast, had 'water' been a flexible designator, 'water'-sentences would have had descriptive truth-conditions: 'water is wet' is true at W_P iff the watery stuff at W_P is wet at W_P . Such truth-conditions are not essentially about any particular kind of stuff but whatever happens to be watery. But the fact that XYZ is watery and wet on a counterfactual Twin Earth does not render Earthly tokens of 'water is wet' true. So it seems that only singular truth-conditions get things right with respect to counterfactual W_P , and thus that 'Aristotle' or 'water' should be modally rigid by (Singular-Rigidity).

I think this line of reasoning is fine as far as it goes. The lesson is that referring terms N cannot have descriptive content ϕ encapsulated in definite descriptions of type (i) - (iv), since that would make N modally flexible with the result that we would get the truth-conditions wrong with respect to counterfactual W_P for tokens of sentences containing N . But what is to prevent the Descriptivist from invoking descriptions of type (vi) as part of

the cluster that make up ϕ ? *Rigidified descriptive truth-conditions* will get things right with respect to such W_P : (15) is true at W_P iff the teacher of Alexander at W_A was fond of dogs at W_P . W_1 is no counterexample as Plato did not teach Alexander at W_A . All the Descriptivist needs is thus a rigidification device which transforms the modally flexible descriptions that constitute the cluster into rigid designators. The notion of rigidity embraced by the Descriptivist should therefore be that of (W_P -Rigidity). One such device is an 'actuality'-operator.³⁹ If 'Aristotle' is short for 'the actual teacher of Alexander', etc., then both N and ϕ are rigid designators by (W_P -Rigidity). 'Aristotle' will pick out at every W_P the individual who taught Alexander at W_A regardless of whether Aristotle or anyone else taught Alexander at W_P . So, 'Aristotle' is still Descriptive in the sense that it picks out its referent at W_P by means of satisfaction of some descriptive condition that also constitutes what someone knows when she understands 'Aristotle'. Or consider our standing example 'water is wet'. If 'water' is rigid by (Singular Rigidity), then an Earthly utterance of 'water is wet' is true at W_P iff H_2O is wet at W_P . But the Descriptivist will insist that 'water' can be rigid by (W_P -Rigidity) and yet have descriptive content given by 'the watery stuff': 'water is wet' is true at W_P iff the watery stuff at W_A is wet at W_P . These rigidified descriptive truth-conditions gets things right with respect to every W_P : Earthly tokens will not be true at a counterfactual Twin Earth, because the watery stuff on Twin Earth - XYZ - is not the watery stuff at W_A - H_2O .

The important point is that, *pace* the Referentialist, there is no straightforward entailment from rigidity to singular content. If N is shorthand for 'the actual ϕ ', then N is Descriptive yet modally rigid by (W_P -Rigidity). So, it is not true that if N is modally rigid, then N is non-Descriptive. It is true that if N is non-Descriptive, then N is modally rigid by (Singular-Rigidity), hence by (W_P -Rigidity), but the Descriptivist will of course deny the antecedent of that conditional. So, it appears that neither (Singular-

³⁹ I shall take the 'actuality'-operator to be an indexical that always refers to the world of the context - even inside the scope of a modal operator. For more on 'actually' see Sec. 4.2.

Rigidity) nor (W_P-Rigidity) can be used against Descriptivism: to assume the former in an argument against Descriptivism is question-begging, to assume the latter, as in the Modal Argument, is harmless once rigidification is available. That is, the Descriptivist may concede that (9) - N might not have been N - is false, but maintain that so is: (10*) N might not have been the actual ϕ .

There are W_P at which N is not ϕ , but if N is ϕ at W_A, then there are no W_P at which N is not ϕ at W_A. Similarly:

(10**) The actual ϕ might not have been the actual ϕ ,

is unambiguously false. There are W_P at which whoever is the unique ϕ at W_A is not ϕ , but there are no W_P at which whoever is the unique ϕ at W_A is not the unique ϕ at W_A. By rigidification ϕ acquires the same modal properties as N, and so there is no basis for setting up a CP-Argument deploying differences in modal properties to individuate content. What the Referentialist must argue is that we have *independent* reasons to accept (Singular-Rigidity) as opposed to (W_P-Rigidity). But if our intuitions about the truth-conditions of sentences containing N with respect to W_P can equally well be accounted for by the truth-conditions of corresponding sentences containing 'actualised ϕ ', then it is hard to see where those reasons should come from.

2.6. The Rigidity Principle

We have seen that the Modal Argument can be blocked by rigidification of the descriptive condition ϕ that, on the Descriptivist account, constitutes the content of a referring term N. This Descriptivist response thus accepts the Referentialist assumption that modal properties are relevant content-individuating properties: if N and ϕ had different modal properties, then N and ϕ would differ in propositional content. But if ϕ is rigidified, then N and ϕ share modal properties, and so N and ϕ do not differ in content. This response thus tacitly endorses the *Rigidity Principle*:

(RP) If N is modally rigid and ϕ modally flexible, then N and ϕ must differ in propositional content.

The Modal Argument crucially hangs on the tenability of (RP). What it says is that if N and ϕ coincided in propositional content, then N and ϕ should be inter-substitutable in all modal contexts *salva veritate*. In other words, if N and ϕ are not co-referring in all W_p , then N and ϕ are not co-propositional. But (9) - N might not have been N - and (10) - N might not have been ϕ - show that substitution fails, and so N and ϕ must differ in content. In particular, since 'If N exists, then N is N' is a *necessary* truth, and *ex hypothesi* N and ϕ have the same content, then 'If N exists, then N is ϕ ' should also be a *necessary* truth. But the truth of (10) shows that it is not, so the hypothesis is false. The argument thus presupposes a *Referentialist Content Assumption*:

(RCA) If N and ϕ have the same propositional content, then 'If N exists, then N is ϕ ' expresses a necessary truth.⁴⁰

What the Modal Argument thus shows is that *if* modal properties like necessity or rigidity are relevant content-individuating properties, *then* if N is modally rigid and ϕ is modally flexible so that 'N is ϕ ' expresses a contingency, then N and ϕ cannot have the

⁴⁰ It is natural to think that (RP) and (RCP) rule that two distinct expressions have the same content only if they are rigid designators of the same thing. But that does not follow. Contraposition on (RP) says that if N and ϕ have the same content, then it cannot be that N is modally rigid and ϕ is modally flexible. So, (RP) does not rule out that, say, the flexible designators 'The individual called 'Cicero'' and 'The individual called 'Tully'' coincide in content. But (RCA) does. 'The individual called 'Tully' is the individual called 'Cicero'' expresses a contingent truth. So, one may think that only true identity statements flanked by rigid designators - Cicero is Tully - are necessary, hence coincide in content. But there are examples of necessary identities between distinct flexible designators: 'The individual denounced by Tully is the individual denounced by Cicero'. Although both descriptions pick out different individuals at different W_p , they will always pick out the same individuals at those W_p .

same content. But no reason has so far been advanced why the Descriptivist must accept the antecedent. Neither (RP) nor (RCA) has been given any substantial support, but the assumption that any difference in modal properties *eo ipso* entails a difference in propositional content is surely in need of such. Descriptivism holds that *meaning and understanding are correlative notions*. The propositional content of N is what a competent speaker knows when she understands N, and to have such knowledge is primarily a matter of grasping certain descriptive conditions that are commonly associated with N. Content is essentially a cognitive notion; it is something that can be internally accessed if asked. But if this is how content is thought of, why should an assumption that stems from a view on which *meaning and modality are correlative notions* be accepted? What someone knows when she understands N is not a function from W_P to referents or extensions as it would be were propositional content simply individuated by its modal properties. We saw in Sec. 2.2 that descriptive content could be *represented* by such a function, namely (DR), but it is *not identical* to (DR). To understand N is not to grasp (DR), although someone who did understand N might have knowledge that issued in correct instances of (DR). Someone who had an impeccable ability to track the right reference of N across W_P could be said to operate (DR), but would not know the content of N if she did not also grasp the associated descriptive conditions. So, given the nature of propositional content on the Descriptivist construal, there is every reason not to accept (RP) or (RCA). To my knowledge, Evans [1996b, p. 182] was the first to contest these principles:

"I agree that sentences containing names embed differently under modal operator than do sentences containing descriptions, but [...] the conclusion which Kripke draws from this fact follows only upon a questionable view of the connection between the content of an utterance and its modal properties."

It is true that a notion of content can be defined such that two expressions have the same content iff they behave similarly in modal contexts. Evans called such a notion a 'proposition', and he distinguished 'proposition' from 'content' as he took content to be a

cognitive notion in the manner just sketched. Evans' point was that just as two sentences can have different contents while expressing the same proposition, two sentences can have the same content whilst expressing different propositions. An example of the former is 'Hesperus is F' and 'Phosphorus is F', and an example of the latter is 'the teacher of Alexander is F' and 'the actual teacher of Alexander is F'.⁴¹ If so, then [1996b, p. 201]:

"...sentences with the same content might embed differently inside the scope of modal operators."

In light of this, it is clear that the Modal Argument has no bite: why should someone who questions the individuation of propositional content by modal properties accept a line of reasoning that presupposes that content is thus individuated? What needs to be shown, and not just assumed, is that (RP) and (RCA) are correct, i.e. that modal rigidity is a relevant property of propositional content, so that Leibniz Law can be used to show that a difference in modal properties always entails a difference in content. The Modal Argument, by itself, provides no reason to think so.⁴²

The Referentialist may rejoin with Kripke [1980, p. 6], cf. Soames [1998, p. 2], that:

"A proper understanding of [(15) Aristotle was fond of dogs] involves an understanding both of the (extensionally correct) conditions under which it is in fact true, *and* of the conditions under which a counterfactual course of history...would be correctly...described by [(15)]."

⁴¹ Note that when one is not engaged in modal discourse, the difference between an utterance of 'the G is F' and 'the actual G is F' seems to be one of conversational implicature, e.g. to stress one's surprise.

⁴² A similar line of resistance can be found in Stanley [1997] and [1999a]. He points out [1997, p. 153], for instance, that not all modal differences can be of relevance to propositional content: there are (metaphysical) W_p at which our words have meanings which differ from the meanings they have at W_A , and surely such W_p are irrelevant.

We agreed that propositional content is truth-conditional so that in order to understand an utterance of:

(15) Aristotle was fond of dogs,

one must know the conditions under which it is true. Now, if (RP) and (RCA) are discarded, as on the proposed view, then an utterance of (15) must have purely descriptive truth-conditions. But we have seen that such truth-conditions are adequate only when it comes to W_A . Non-rigidified descriptive truth-conditions get things wrong with respect to counterfactual W_P . But, as Kripke says, an understanding of (15) cannot consist merely in a grasp of the conditions under which such an utterance is true at W_A . (15) and

(16) The teacher of Alexander was fond of dogs,

have the same truth-conditions with respect to W_A - the same W_A -truth-conditions - in the sense that Aristotle's affection for dogs at W_A would make utterances of both true at W_A .⁴³ But, the objection goes, that cannot be right, since proper understanding must also involve knowledge that the conditions under which utterances of (15) and (16) are true at various counterfactual W_P are different, i.e. that they have different W_P -truth-conditions. Someone who thinks that (15) is true at W_P iff the teacher of Alexander at W_P is fond of dogs at W_P has an incorrect understanding of (15), and someone who is agnostic about the W_P -truth-conditions of (15) has only a partial understanding of (15). The point is that if propositional content is truth-conditional, as we suppose, then since utterances of (15) and (16) differ in truth-value in some W_P , their truth-conditions must differ at some W_P , so their content must differ. In other words, if (W_P -Rigidity) is accepted, then propositional content must be individuated by a modal notion of truth-conditions. But understanding is knowledge of propositional content. So, mere knowledge that the very same state of affairs make both

⁴³ Someone might say that they have *different* W_A -truth-conditions: (15) is true iff Aristotle was fond of dogs, whereas (16) is true iff the teacher of Alexander was fond of dogs. The point is merely that the very same state of affairs - Aristotle's affection for dogs at W_A - make both true at W_A ; not that those state of affairs cannot be given different descriptions.

(15) and (16) true at W_A falls short of constituting everything a speaker has to know if she is to have full understanding of them.

Or take another example. Utterances of the sentences 'Animals with hearts have kidneys' and 'Animals with hearts have hearts' coincide W_A -truth-conditions, but clearly have different contents. Co-extensionality at W_A is too coarse-grained a criterion for content-individuation. One must also know various counterfactual conditions under which such utterances would be true or false, e.g. that some animals could have had hearts without kidneys but not hearts without hearts. If all a speaker knows is that 'animals with kidneys' and 'animals with hearts' are co-extensional at W_A , without appreciating their distinct modal properties, then she will be ignorant of essential aspects of their correct assertoric usage. So, the Referentialist objection is that sameness of W_A -truth-conditions cannot be what individuates propositional content without remainder.

The question is whether this point - essentially that 'truth-condition' is a modal notion - vindicates Referentialism. Kripke [1980] clearly thought so. After the quotation I gave above he goes on to define modal rigidity in terms of (Singular-Rigidity).⁴⁴ It is of course true that the singular truth-conditions - 'Aristotle was fond of dogs' is true iff Aristotle (the man himself) was fond of dogs - give the right conditions under which (15) is true both at W_A and at counterfactual W_P . So, if a speaker grasps those singular truth-conditions, then she knows both its W_A -truth-conditions and its W_P -truth-conditions. Nevertheless, once rigidification is available, the Descriptivist can easily accommodate the claim that full understanding requires knowledge of W_A -truth-conditions as well as knowledge of W_P -truth-conditions without endorsing singular truth-conditions. There is no need to have knowledge of reference since knowledge of - rigidified - descriptive content will suffice. Consider what Dummett [1991, p. 47], says about truth-conditions:

"To grasp the content of an assertion, one needs to know only what possibilities it rules out, or, positively expressed, under what conditions it is correct."

⁴⁴ See the other quotation from Kripke [1980, p. 6] I gave in Sec. 2.5.

By *content* Dummett understands what an assertoric utterance of an atomic, i.e. unmodalized, sentence expresses. If there were widespread consensus amongst competent speakers that 'Aristotle' picked out whoever was the teacher of Alexander at W_A , then (15) and (16) would express the same contents.⁴⁵ 'Aristotle' and 'the teacher of Alexander' would determine their referents at W_A in the same way, and so (15) and (16) would say the same thing about how W_A is. But how can that be if what someone knows when she grasps the content of an assertion is which *possible* state of affairs would render it correct and which incorrect - different possible state of affairs make (15) and (16) true? Well, it is clear from the context in which this quotation occurs,⁴⁶ that Dummett does not take 'possible' to mean *counterfactual possibility*. The claim is rather that the same possible circumstances at W_A would make (15) and (16) true. It is *impossible* to conceive of *actual* circumstances under which the one, but not the other, is true. At W_A it is Aristotle's fondness for dogs which makes (15) and (16) true. But suppose that Plato had been the teacher of Alexander at W_A . Then if 'Aristotle' simply refers to whoever 'the teacher of Alexander' picks out at W_A , then 'Aristotle' would refer to Plato. If Plato moreover had been fond of dogs, then (15) would have been true at W_A due to Plato's fondness for dogs. But the very same state of affairs would have made (16) true. No matter how W_A had turned out, (15) is true at W_A iff (16) is true at W_A . So, we may say that someone who knows the *purely descriptive content* of 'Aristotle' knows the W_A -truth-conditions of sentences containing 'Aristotle': she knows not only what makes them true in W_A , but also what would make them true had W_A been different in various ways.⁴⁷

⁴⁵ The example is in fact not very plausible and Dummett [1981, p. 562] rightly complains that it has a prejudice against Descriptivism. I shall discuss Evans' example, which is very similar to Dummett's, at some length in Sec. 4.3.

⁴⁶ [1991, pp. 47-8], cf. also Dummett [1973, pp. 134-5; 1981, pp. 564-8].

⁴⁷ I shall get back to this point in Chapter 4.

What about counterfactual possibilities then? Well, just as Evans distinguished between content and proposition, Dummett makes a distinction between (assertoric) *content* and *ingredient sense*: the ingredient sense of a sentence is the contribution the sentence yields to determine the content of more complex sentences of which it is a sub-sentence.⁴⁸ Just as on Evans' account, (15) and (16) coincide in content, but differ in ingredient sense/propositions. They *coincide in W_A -truth-conditions* in the sense just described, but they *differ in W_P -truth-conditions* in that they embed differently inside the scope of a modal operator. So, although both Evans and Dummett would dispute (RP) and (RCA) as principles pertaining to what they call 'content', (RP) and (RCA) are sustained by what they call 'proposition' or 'ingredient sense'. The crucial point is that 'proposition' or 'ingredient sense' need not be notions of singular content. This is where rigidification kicks in. Sentences containing 'Aristotle' and 'the teacher of Alexander' have the same W_A -truth-conditions, but sentences containing 'Aristotle' and 'the actual teacher of Alexander' have the same W_P -truth-conditions. That was how we responded to the Modal Argument: 'Aristotle' and 'the actual teacher of Alexander' are intersubstitutable *salva veritate* in modal contexts. So, Kripke is right that full understanding requires knowledge of W_P -truth-conditions, but he is wrong that only knowledge of singular truth-conditions will fit the bill. All it takes is knowledge of *rigidified descriptive truth-conditions*, i.e. that 'Aristotle' is a rigid designator of whoever taught Alexander at W_A . Someone who knew that would know that, say, 'Aristotle might not have been the teacher of Alexander' is true at W_P iff whoever taught Alexander at W_A did not teach Alexander at W_P . No knowledge of reference is called for.

The picture I have in mind can be illustrated by Dummett's distinction between two grades of understanding.⁴⁹ A speaker who grasps the purely *descriptive truth-conditions* of

⁴⁸ Cf. Dummett [1973, pp. 446-7; 1981, pp. 558-73; 1991, p. 48]. See also Forbes [1989] for a distinction on much the same lines between the sense of an expression and its cognitive significance.

⁴⁹ Op. cit.

an atomic sentence has knowledge of its W_A -truth-conditions. She has a *grade-one* understanding of that sentence. As Dummett suggests, it is presumably feasible to have a mere grade-one understanding, that is, to know under which conditions an atomic sentence is true at W_A without knowing the conditions under which it is true at various counterfactual W_P . Perhaps she knows how the reference of a proper name occurring in that sentence is fixed at W_A , but is unaware that the name is a rigid designator, and so is unable to determine its reference with respect to counterfactual W_P . Clearly, such a speaker's understanding would be essentially incomplete in that there would be an important aspect of the assertoric use of that sentence which the speaker did not master, namely how to determine the truth-value of complex, modalized sentences containing it. If, however, a speaker knows the *rigidified descriptive truth-conditions* of a sentence, then she has knowledge of such W_P -truth-conditions. She has a *grade-two understanding* in that she knows the conditions under which that sentence is true at counterfactual W_P . To have a grade-two understanding of a sentence is to know the modal properties of the contained expressions. But to have knowledge of such properties need not involve knowledge of reference. Given that modal rigidity is cashed out by (W_P -Rigidity), one can know that 'Aristotle' is a rigid designator and yet lack knowledge of its referent. So, the Referentialist is right that one cannot embrace the modal rigidity of referring terms without at the same time accepting a notion of propositional content which is individuated by that property. But the Referentialist is wrong that only singular content is thus individuated. Singular terms can have descriptive content if only conferred upon them by rigidified expressions.

Let me sum up. The Modal Argument should only carry conviction against a form of Descriptivism according to which a referring term N is modally rigid yet its only propositional content is given by modally flexible definite descriptions ϕ . Since the Modal Argument assumes that N is a rigid designator, a form of traditional Descriptivism which denies that assumption can also consistently be held although only with highly counter-intuitive consequences. There are, however, also more plausible ways to block the Modal

Argument. One is to meet the underlying constraints - (RP) and (RCA) - about content-individuation. If the descriptive content of a rigid designator N were given by rigidified ϕ , then Leibniz Law could not be used to show that N and ϕ must differ in content since there would then be no difference in modal properties between N and the content-conferring ϕ . The Descriptivist could even carve up descriptive content into a purely descriptive component and a rigidified component where only the latter would be subject to (RP) and (RCA). Descriptive content is cognitive in nature and not individuated by its modal properties. It is thus possible to have - a grade-one - understanding of an atomic sentence without knowing how that sentence would embed inside complex, modalized sentences. In any case, the point is that there is no easy move from modal rigidity to lack of descriptive content. What we learn from the Modal Argument is not that referring terms are purely Referential, but that they cannot be purely Descriptive. It has taught us that we must supplement the cluster of descriptive properties with a rigidification device such that those properties only determine reference at W_A . It has not taught us that reference does not go by associated properties, but rather *which* properties it goes by.

Chapter 3. The Intentional Argument

3.1. Substitutivity

We have defined Descriptivism as the view that referring terms N - proper names and natural kind terms - have associated with them a cluster of descriptive conditions ϕ which is both what someone knows when she understands N and what determines the reference of N at W_P - the referent of N being, if anything, all and only objects o which uniquely satisfy ϕ at W_P . In contrast, Referentialism is the view that N has no descriptive content. The reference of N may have been fixed by satisfaction of ϕ to be o , but since N is a rigid designator while ϕ a flexible designator, ϕ cannot be what determines the reference of N at W_P , so ϕ cannot be what constitutes the propositional content of N . The content of N is singular in being provided by o such that someone who understands N must know of o that N refers to o .

What supports Referentialism are thus considerations about the behaviour of N in modal discourse; whence, the Modal Argument. The Referentialist purports to show by an instance of the CP-Argument that a difference in modal properties between N and ϕ must imply a difference in propositional content, hence that N is non-Descriptival. We have by now reason to think that this Argument does not accomplish what it sets out to prove. It rests on two assumptions: (i) N is modally rigid, and (ii) a modally rigid N never has the same content as a modally flexible ϕ . Old-style Descriptivism would flatly deny (i), whereas a more plausible new-style Descriptivism would endorse (i) and invoke some rigidification strategy to account for (ii). Alternatively, one could refuse to accept (ii) by allowing for a bifurcation of propositional content. In any case, the pivotal Descriptivist tenet - that N has associated descriptive content - has not been jeopardised by the Modal Argument. The question is now: if the main Argument against Descriptivism fails in its

intent, are there any arguments that compel endorsement of this Descriptivist tenet? In this Chapter I shall motivate an affirmative answer. First some more background.

According to Referentialism, a referring term *N* is not only modally rigid, but also lacks descriptive content; indeed the truth of the former follows from the truth of the latter. Purely Referential terms are first and foremost ordinary proper names, but it is intrinsic to the view that also natural kind terms are Referential.¹ The propositional content of *N* is therefore supplied by the object *o*, if any, to which *N* refers. Whatever other 'content' may be conveyed by use of *N* is of a non-semantic nature. But if the semantic function of *N* is merely to serve as a referring device, then it should be safe to substitute *N* for any co-referring purely Referential term *M* in sentences in which *N* occurs without change in truth-value or proposition expressed by those sentences; if *N* and *M* are co-referring, then they are also co-propositional. This is certainly true when we consider only atomic sentences, i.e. singular terms under predication and outside the scope of any modal or other intentional operators. Thus suppose that 'Hesperus' is purely Referential, and is substituted in:

(1) Hesperus is the evening star,

by the co-referring purely Referential name 'Phosphorus':

(2) Phosphorus is the evening star.

It is clear that utterances of (1) and (2) express the same singular proposition: <Venus; being the evening star>; a proposition consisting of the planet Venus and the property of being the star that is visible in the evening at such-and-such a position. But it is not only under predication that co-referring purely Referential terms are intersubstitutable without change in truth-value or content. The fact that they are rigid designators guarantees that similar changes in modalized sentences will also always preserve both truth and content. Thus utterances of:

¹ In [1980, pp. 127-8, 134-5; 1994, p. 374], Kripke explicitly modelled manifest natural kind terms on ordinary proper names, and did seem to hold the view that both were purely Referential, or 'Millian' as he preferred, although with some reservations as we shall see.

(3) Hesperus might not have been Hesperus,

and

(4) Hesperus might not have been Phosphorus,

are both false, because they both say of the planet Venus that it might not have been self-identical. It is a consequence of Referentialism that two purely Referential terms - N and M - are co-referring at W_A iff they are co-referring at all W_P , and that they are co-referring at all W_P iff they have the same singular content. So, N and M are co-referring at W_A iff they have the same singular content. The first bi-conditional follows from the fact that both N and M are modally rigid and explains why N and M are intersubstitutable in modal contexts *salva veritate*, and the second bi-conditional follows from the fact that N and M lack descriptive content and explains why N and M are intersubstitutable in modal contexts *salva significatione*. It is thus part of Referentialism that N and M have the same content iff they have the same modal properties. That explains why it was assumed in the Modal Argument that if N and ϕ have different modal properties, then they must differ in content, i.e. why the Referentialist endorses (RP) and (RCA). Here is Kripke [1994, p. 353]:

"If a strict Millian view is correct, and the linguistic function of a proper name is completely exhausted by the fact that it names its bearer, it would appear that proper names of the same thing are everywhere interchangeable not only *salva veritate* but even *salva significatione*: the proposition expressed by a sentence should remain the same no matter what name of the object it uses."

This suggests that Referentialism is committed to the following *Principle of Substitutivity*:

(PS) If $E(\alpha)$ is a sentence containing a referring term α , then substituting α by a referring term β does not change the truth-value of $E(\alpha)$, if α and β have the same reference.

Note the resemblance between (PS) and Leibniz Law: $[\alpha = \beta \ \& \ E(\alpha)] \rightarrow E(\beta)$. Since Leibniz Law is firmly entrenched, it would seem as if we are committed to (PS) across the

board. We have seen that (PS) must find application on atomic and modal sentences, but are any linguistic contexts exempted? Consider a - slightly modified - famous example²:

(5) Hesperus was so-called because of its evening visibility.

It goes without saying that substituting 'Hesperus' for 'Phosphorus' in (5) results in change in truth-value. Venus was called 'Phosphorus' because of its morning visibility. So, (PS) fails in contexts which involve quotational devices, e.g. mention of words. If we follow Quine and call contexts in which (PS) holds (referentially) *transparent* contexts and contexts in which it does not (referentially) *opaque*, then it should be indisputable that quotational contexts are opaque. It is worth emphasising why (PS) fails in (5). As 'Hesperus' occurs in (5), it does not merely have its customary reference as it does in (1), (3) and (4). In (5) 'Hesperus' refers to the name 'Hesperus' as well as the planet Venus. In the terminology, 'Hesperus' occurs in opaque position. This suggests that the reference of a term is sensitive to the linguistic context in which it occurs, i.e. that 'reference' in (PS) be qualified. Here is an amended Principle of Substitutivity which, at least on the face of it, admits of no exceptions:

(PS*) If $E(\alpha)$ is a sentence containing a referring term α , then substituting α by a referring term β does not change the truth-value of $E(\alpha)$, if the reference of α in $E(\alpha)$ is the same as the reference of β in $E(\beta)$.³

(5) is no counterexample to (PS*): 'Hesperus' and 'Phosphorus' are not co-referential, since in the quotational context E they refer *inter alia* to different proper names. As Quine pointed out, there are also counterexamples to (PS) when the context is modal. Famously,

(6) It is necessary that $9 > 7$,

is true, but substituting '9' for the - at W_A - co-referring 'the number of planets',

(7) It is necessary that the number of planets > 7 ,

² Quine [1960, pp. 141-156; 1980, pp. 139-59].

³ The idea of contextualising Substitutivity owes much to Forbes [1989, p. 121].

fails to preserve truth. (PS*), however, can handle (6) and (7), since for α and β to be co-referring in modal contexts requires that α and β are co-referring at *all* W_P . In other words, when the context E is modal, and α and β differ in modal properties, then it is possible that ' $\alpha = \beta$ ' is true, while ' $E(\alpha) \leftrightarrow E(\beta)$ ' is false. So, the failure of substitution is due to the fact that while '9' is a rigid designator, 'the number of planets' is not.⁴

Are there any linguistic contexts where (PS*) breaks down? If we turn to *intentional contexts* and in particular ascriptions of De Dicto propositional attitudes, we have firm intuitions that it does. It is, for instance, part of the Superman legend that Lois Lane never realises that Superman is, in fact, identical to Clark Kent. Our inclination is to think that since Lois Lane is a rational speaker who sincerely and on reflection assents to:

(8) Superman flies,

but, given her state of information, reflectively and honestly dissents from:

(9) Clark Kent flies,

it is true to say that:

(10) Lois Lane believes that Superman flies,

but false to say that:

(11) Lois Lane believes that Clark Kent flies.

The problem is, however, that if both 'Superman' and 'Clark Kent' are purely Referential, then (PS*) should allow us to infer (11) from (10). If the propositional content of 'Superman' is exhausted by its referent, then since 'Superman' and 'Clark Kent' have the same referent, they should also have the same propositional content; hence any two sentences - (8) and (9) - which differ only in that the one contains 'Superman' whereas the

⁴ Note that if 'the number of planets' is either rigidified or read as taking wide scope with respect to the modal operator - the number of planets is such that necessarily it is larger than 7 - then the substitution preserves truth. As is well-known, Smullyan [1948] took this to refute Quine's startling claim [op. cit.] that modal logic failed to sustain Substitutivity and thereby violated Leibniz Law.

other contains 'Clark Kent' should express the same proposition. This provides sufficient background for setting up a CP-Argument against Referentialism.

The *Intentional Argument* goes like this: suppose 'Superman' and 'Clark Kent' coincide in propositional content by the lights of Referentialism, i.e. are co-referring purely Referential terms. It follows by (*Compositionality of Content*)⁵ that the contents expressed by (8) and (9) should be identical. But if two contents are identical, then, by Leibniz Law, they should have the same content-individuating properties. In particular, they should have the same *cognitive properties* where this is understood to mean that the sentences that express those contents must embed similarly inside the scope of an attitude-operator. Hence, (10) and (11) should also express the same contents. Yet, we agreed, (10) is true, while (11) is false. But a difference in truth-value entails a difference in content. So, (10) and (11) must differ in content, hence, so must (8) and (9), and therefore also 'Superman' and 'Clark Kent'. And since nothing hangs on the particular example - same stories could be told about Hesperus/Phosphorus, Cicero/Tully, etc. - the conclusion generalises to all purely Referential terms.

In other words, were Referentialism true, (PS*) should license intersubstitutions of co-referring purely Referential terms in intentional contexts *salva veritate*, i.e. such contexts should be transparent. On this view, co-referentiality entails co-propositionality. But this manifestly jars with best intuitions about the truth-values of the resulting sentences. A term occurring inside the scope of a 'believe that'-operator cannot in general be supplanted by a co-referring term without change in truth-value: it occurs in opaque position. So, co-referentiality of referring terms is - the thought goes - too coarse-grained a criterion for sameness of their propositional contents.

⁵ Along the lines of: the propositional content of a complete complex expression is uniquely determined by the propositional contents of its component expressions and the way in which they are combined; cf. also Sec. 2.4.

It is natural to think of a *proposition* as the content of a propositional attitude, as the object of belief. Suppose therefore we accept the *Relational Principle of Belief*:

(RPB): S believes that P states a relation between S and the proposition expressed by 'P'.

Then, on the Referentialist account, (10) states that S stands in the belief-relation to the singular proposition about Superman that he flies: $\langle \text{Superman}; \text{fly-hood} \rangle$. On this account, remember, all there is to the propositional content of 'Superman' is its bearer. So, if Lois Lane believes the proposition expressed by the embedded sentence 'Superman flies', she must stand in a belief-relation to a singular proposition individuated by Superman, i.e. Clark Kent, and the property of flying. (11) thus follows from (10), because (11) states that the same believer stands in the same cognitive relation to the same singular proposition. The mere fact that 'Superman' and 'Clark Kent' are co-referential guarantees that their singular contents coincide, hence that the embedded sentences in (10) and (11) express identical singular propositions. Moreover, if it is true that Lois Lane believes that Superman flies, then there must be a proposition she believes which is singular with respect to Superman such that had he not existed that proposition would not have existed. It is thus always safe to *existentially generalise*⁶ - infer $(\exists x)Fx$ from Fa - on a true De Dicto belief ascription like (10):

(12) There is someone such that Lois Lane believes that he flies.

In other words, whenever a structurally De Dicto belief ascription - an ascription in which the singular term takes narrow scope with respect to the cognitive operator - is true, its structurally De Re counterpart is also true:

(13) Lois Lane believes of Superman that he flies.

⁶ As Quine remarked [op. cit.] existential generalisation fails in opaque contexts. For instance, from the fact that S believes that Hesperus was so-called because of its evening visibility, one cannot infer that there is something such that S believes that it was called 'Hesperus' because of its evening visibility. S may just believe that if 'Hesperus' names anything, then only in virtue of that thing being visible in the evening.

The fact that interchanging 'Superman' for 'Clark Kent' in (13) preserves truth-value shows that 'Superman' occurs in transparent position in (13). More vividly, (13) is semantically equivalent to

(14) Superman is such that Lois Lane believes that he flies,

where 'Superman' appears outside the scope of the 'believe that' operator and so is clearly in transparent position, hence subject to (PS*). Conversely, whenever a De Re belief ascription like (13) is true, its structurally De Dicto counterpart in (10) is also true: to believe of *a* that it is *F* is just to believe the singular proposition about *a* that it is *F*. Or formally, ' $B(S, Fa)$ ' is true iff ' $\exists x [x = a \ \& \ B(S, Fx)]$ ' is true, where ' $B(x, y)$ ' reads: *x* bears the belief-relation *B* to *y*. So, on the Referentialist account there are only syntactic differences between (10), (13) and (14). All Referential terms, whether in De Dicto or De Re belief ascriptions, occur in transparent position which is to say that all true syntactic De Dicto belief ascriptions containing purely Referential terms are semantically equivalent to their syntactic De Re counterparts. Were Referentialism true, referring terms would thus not exhibit any semantic De Re - De Dicto distinction in intentional contexts.⁷ What the Intentional Argument purports to show is that doxastic contexts should be governed by such a distinction, i.e. that it should allow for opaque occurrences of referring terms and so that (PS*) should not permit intersubstitution of all co-referring terms. To think otherwise is to disregard speakers' linguistic behaviour and intuitions.

3.2. Semantics or Pragmatics

Prima facie implausible as it seems, Referentialism has nevertheless been defended with much ingenuity and persistence.⁸ Let me briefly canvass how the defence goes. The idea is,

⁷ In other words, purely Referential terms would be scopeless just as we saw that they were in modal contexts.

⁸ Notably in Salmon [1986; 1989], and Soames [1987a; 1987b].

roughly speaking, to explain away the intuitive readings of (10) and (11) as not genuinely pertaining to their semantics, but to what they pragmatically implicate - hence *the Implicature Theory*. What is called for is some persuasive story why we have such strong intuitions that (10) and (11) can be given opaque readings. The Implicature Theory is helpful in this respect in that it explains the appearance of opacity as due to, not a difference in semantics, but in conversational implicature. The Theory adheres to (RPB) with the modification that a belief sentence 'S believes that P' be analysed as a three-place relation holding among believers, singular propositions and something else:

(15) $(\exists x)$ [S grasps P by means of x & B (S, P, x)], where x is, say, the disposition to assent when the proposition P is taken in a certain way. But an utterance of 'S believes that P' communicates not only that *there is a way* in which S takes the proposition she believes, but also *which way* that is, namely that S is presented with P through the sentence in the that-clause 'P':

(16) B [S, that P, x(S, 'P')],

where $x(S, 'P')$ is the way S would take the proposition expressed by the sentence 'P', were it presented to S through 'P'.⁹ On this view, (10) and (11) receive identical truth-values, since the same singular proposition is believed in different linguistic guises. The mistaken intuition that (11) is false, while (10) is true, is due to the fact that while (10) has the *true pragmatic implicature* that Lois Lane is disposed to assent to 'Superman flies', (11) has the *false pragmatic implicature* that Lois Lane is disposed to assent to 'Clark Kent flies'. (10) and (11) communicate different things, but - strictly speaking - say the same thing. Given her state of information, Lois Lane is disposed to dissent from 'Clark Kent flies', or to assent to 'Clark Kent does not fly'. Had Lois Lane been in a different state of information which comprised knowledge of their identity, she would have been disposed to assent to 'Clark Kent is Superman', and so also been disposed to assent to 'Clark Kent flies' given that she is disposed to assent to 'Superman flies'. She would thus have to withdraw her

⁹ (15) and (16) are explicit in Salmon [1986].

assent to 'Clark Kent does not fly', since she could not rationally assent to both 'Clark Kent flies' and 'Clark Kent does not fly'. But on the assumption that Lois Lane fails to realise their identity, she can rationally believe and disbelieve that Superman flies. In our notation:

(17) B [Lois Lane, that Superman flies, $x(\text{Lois Lane, 'Superman flies'})$],

is true, while:

(18) B [Lois Lane, that Superman flies, $x(\text{Lois Lane, 'Clark Kent flies'})$],

is false. When beliefs are reported, one should strive "...to remain faithful to the words of the speaker unless there is reason to deviate."¹⁰ One is constrained to ascribe beliefs whose content-clauses contain only sentences that S would assent to were she presented with them. Otherwise an utterance of a belief sentence is liable to mislead a hearer. It is clear that (11) violates this *Maxim of Faithfulness* and is thus *inappropriate*. That, however, is not to say that it is *false*. It is literally true that Lois Lane believes that Clark Kent flies, since she does stand in the belief-relation to the singular proposition expressed by 'Clark Kent flies'. It is just that that ascription falsely implicates that it is in virtue of accepting 'Clark Kent flies' that she believes that proposition. The contention is thus that even if competent speakers are highly sensitive to the informational content conveyed by utterances of (10) and (11) respectively, they are equally unreliable in sorting out what is *semantically encoded* and what is *pragmatically imparted*.¹¹ They are, in particular, prone

¹⁰ Soames [1987a, p. 119].

¹¹ That - untutored - speakers tend to confuse implicated information with truth-conditional content is familiar from 'Jane became pregnant and she got married' which seems to imply that the former happened before the latter. But it is no part of the truth-conditional content of 'and' to encode a temporal priority between its conjuncts. An utterance of this sentence would be true even if Jane got pregnant after her wedding; the example is Salmon's [1989, pp. 252-3]. If, however, that sentence features as antecedent in 'if Jane became pregnant and got married, her father must be unhappy', what seemed to be temporally implicated must now be semantically encoded; cf. Recanati [1997, pp. 342-3].

to think that informational content at the level of conversational implicature goes into determining the truth-conditions of the reported attitudes.¹²

One worry immediately springs to mind: it is not hard to see how the implicature could be false once it is there, but rather how it could as much as arise in the first place. Bear in mind the point from Sec. 3.1 that, following Referentialism, even though both of (10) and (11) are syntactically De Dicto, they can be given nothing but semantically De Re readings, indeed the entire De Dicto/De Re distinction seems to vanish at the level of semantics. Despite the structural difference, (10) is semantically equivalent to its syntactically De Re counterpart (13). But how could (13) give rise to any implicature? No doubt, (13) implies that Lois Lane has some way of referring to Superman, and presumably she will be disposed to assent to a sentence containing a term in subject position that refers to Superman - ' ψ flies'. But it is no part of the functioning of 'Superman' in (13) to indicate which way of referring that is. We could equally well have reported the very same De Re belief in a different linguistic guise. But if it is no part of De Re belief ascriptions that one rather than another linguistic mode of presentation of reference is implicated, then how can it be part of De Dicto belief ascriptions when they are claimed to be semantically equivalent to their De Re counterparts? It seems that only if (10) were given an opaque reading could it be accounted for how an utterance could pragmatically impart that Lois Lane was disposed to assent to 'Superman flies', for only then need one take into account the linguistic mode under which the believed proposition is presented to her. In embracing only transparent readings, the Implicature Theory has simply divested itself of the

¹² Note how easily the Implicature Theory can handle Frege's [1994b] original puzzle about how ' $a = a$ ' and ' $a = b$ ' can differ in *cognitive value* if ' a ' and ' b ' are purely Referential. The two sentences convey different information - only the first is a priori knowable - but they have the same singular truth-conditions. What one must do is prefix to those sentences an appropriate intentional operator, e.g. 'It is a priori knowable that' or 'S believes that', such that the complex sentences receive different truth-values. Only that way can one hope to argue via (Compositionality of Content) that the difference between the two atomic sentences is semantically relevant.

possibility of explaining, not only why one implicature arises rather than another, but even how they could as much as get off the ground in the first place.¹³

Now, maybe the Referentialist has the resources to patch up the Implicature Theory and hit upon some credible story as to why embedded expressions that occur in transparent positions nevertheless can carry certain pragmatic implicatures. It is still worth emphasising that if we allow for the same expressions to occur in opaque positions, then there is an obvious explanation as to why those implicatures can arise. So, not only can a view that treats doxastic contexts as opaque accommodate speakers' intuitions about the truth-conditions of belief ascriptions, it can also easily explain, by appeal to that very opacity, why such ascriptions have a certain pragmatics. Yet distorted intuitions and explanatory power is one thing, a knock-down argument another.

Before we address the question of whether more troublesome objections can be levelled against Referentialism, there is a key question which I have left hanging in the air: does this response on behalf of Referentialism accept the assumption, underlying the Intentional Argument, that the *cognitive properties* of referring terms are *relevant content-individuating properties*? Reflection on how Referentialism was characterised in Sec. 2.3 certainly suggests that the contents of referring terms should not be individuated intentionally, i.e. by their behaviour in belief contexts. Two referring terms have the same propositional content iff they are co-referring and lack descriptive content. If 'Superman' and 'Clark Kent' are co-referring, then supplanting one for the other in more complex expressions should neither result in change of truth-value nor have an impact on the proposition expressed - even if intentional operators are prefixed to those expressions. Nevertheless, the answer depends on what is understood by 'cognitive property'.¹⁴ It is clear that the Intentional Argument rests on the following *Descriptivist Content Assumption*:

¹³ Cf. also McKinsey [1998, pp. 6-10], who argues along the same lines with respect to indexicals.

¹⁴ There are special problems with second-order properties: even if N and M are synonyms, S could competently believe that S* believes that N is F without believing that S* believes that M is F; or if not

(DCA) If 'P' and 'Q' have the same propositional content, then a fully competent speaker S believes that P iff S believes that Q,

where 'P' and 'Q' are atomic sentences free of intentional and modal vocabulary. But (DCA) is not disputed by the Implicature Theory. (10) and (11) contain (8) and (9) respectively within the scope of a believe-that operator yet both are true precisely because they state that Lois Lane believes the same proposition that (8) and (9) express. One may of course say that since the Theory bites the bullet and assigns the same truth-values to (10) and (11), it reduces the role of cognitive properties as merely pertaining to pragmatics. In that sense the Theory does contest an assumption concerning content individuation underlying the Intentional Argument:

(DCA*) If 'P' and 'Q' have the same propositional content, then a fully competent speaker S is disposed to assent to 'P' iff S is disposed to assent to 'Q'.

On the Implicature Theory, (DCA*) is rejected. The propositional content of (8) is identical to that of (9), but Lois Lane is only disposed to assent to (8), since she is not presented with that content through (9). The point is merely that, as far as semantics is concerned, 'Superman' and 'Clark Kent' have the same cognitive properties in the sense that they are intersubstitutable in belief-contexts without change in truth-value. In order to properly question (DCA), it would have to be conceded that 'Superman' and 'Clark Kent' were not everywhere interchangeable *salva veritate* yet denied that this would make a difference in propositional content. Similarly, we saw that the Descriptivist might deny (RP) and (RCA) underlying the Modal Argument: although replacement of co-referring terms in modal contexts might fail to preserve truth-value, they could nevertheless coincide in

believe then doubt. In order to avoid such *hyper-intentional* contexts we should just focus on first-order beliefs held by fully competent speakers.

propositional content. This strategy has, to my knowledge, not been adopted by any form of Referentialism.¹⁵

3.3. The Intentional Argument

We have seen that Referentialism is committed to the universal application of (PS*) in intentional contexts: any two co-referring purely Referential terms are interchangeable in such contexts *salva veritate*. We are now in a position to fortify our initial Intentional Argument using (PS*) in conjunction with two platitudes governing our practice of reporting beliefs. The first is *Disquotation* which connects disposition to assent with belief:

(D) A competent speaker S is, on reflection, sincerely disposed to assent to 'P' iff S believes that P,

where 'P' ranges over appropriate assertoric sentences lacking indexical devices and ambiguities. By 'competent', we assume that S uses 'P' in accordance with standard usage, by 'sincerely' that S is not deceitful, and by 'on reflection' that S is attentive and without mental defects. The rationale behind (D) is that to utter an assertoric sentence 'P' is, *ceteris paribus*, to assert that P and that to assert that P is to express the belief that P. Conversely, to believe that P is, *ceteris paribus*, to be disposed to assert that P by making an assertoric

¹⁵ I suppose (*Compositionality of Content*) would have to be abandoned. If 'Superman' and 'Clark Kent' coincide in content, and that is the only difference between (10) and (11), then, by (*Compositionality of Content*), (10) and (11) should also coincide in content. The same need not follow on the version of Descriptivism which accepts modal rigidity but denies (RP), since this view, remember, embraced a notion of propositional content - Evans' 'proposition' and Dummett's 'ingredient sense' - which was individuated by (RP). So, this other notion could explain why (*Compositionality of Content*) was not jeopardised; indeed it was explicitly intended to by Dummett [1981].

utterance of 'P' in appropriate circumstances.¹⁶ The second is *Consistency* which links rationality with absence of logically contradictory beliefs:

(C) A fully rational speaker S cannot reflectively and occurrently believe that a is F and that a is not F.

There are undoubtedly cases where speakers rationally believe things they are not reflectively aware of. Psycho-analysis abounds with agents who endure cognitive illusions, repressed prejudices, self-deceptions, etc. Put them aside. (C) is only concerned with occurrent and reflectively accessible beliefs held by speakers who are as rational as anyone gets. We can now use (D), (C) and (PS*) to set up a *Strengthened Intentional Argument*:

- | | |
|--|-------|
| (i) S is fully rational | ass. |
| (ii) $a = b$ | ass. |
| (iii) S assents to 'a is F' and S assents to 'b is not F' | ass. |
| (iv) S believes that a is F and S believes that b is not F | (D) |
| (v) S believes that a is F and S believes that a is not F | (PS*) |
| (vi) S has contradictory beliefs, so S is not fully rational | (C) |

Assumptions (i) - (iii) are uncontested, (D) and (C) seem warranted, and the reasoning is fine. But since (i) contradicts (vi), something must give way. Clearly, it is the application of our Substitutivity Principle:

(PS*) If $E(\alpha)$ is a sentence containing a referring term α , then substituting α by a referring term β does not change the truth-value of $E(\alpha)$, if the reference of α in $E(\alpha)$ is the same as the reference of β in $E(\beta)$,

to move from (iv) to (v) which generates the trouble. But according to Referentialism, (PS*) should license us to infer (v) from (iv) on the assumption in (ii) that $a = b$. This view

¹⁶ We also assume that S speaks the same language as the speaker who attributes the beliefs to S, i.e. that some homophonic translation principle holds, and that the speaker is not reticent - we occasionally believe things we never get the opportunity, or have the courage, to express; cf. Kripke [1994, pp. 360-1, 364, 375].

has it, remember, that co-referring purely Referential terms a and b are intersubstitutable in intentional contexts like (iv) and (v) without change in truth-value.

Note that what (C) rules out, on pain of irrationality, is that S can reflectively and simultaneously hold contradictory De Dicto beliefs. There is no inconsistency in holding similar De Re beliefs. Lois Lane, for instance, may both believe of Superman that he flies and that he does not fly. Substitute throughout and this follows from (iv). The reason why those beliefs do not impugn her rationality is that their contradictory nature cannot be resolved by reflection; she must realise, by whatever empirical means, that Superman is Clark Kent before she could be moved to withdraw her assent to 'Clark Kent does not fly'. Not so for De Dicto beliefs. If Lois Lane, on reflection, did occurrently believe that Clark Kent flies and that Clark Kent does not fly, then she would incur criticism for being inconsistent. The problem is that the conjunction of (D) and (PS*) validates just that in (v) yet she is not inconsistent.

What can be said in response on behalf of Referentialism? There is no way of circumventing the conclusion in (v) that S holds *logically contradictory beliefs* which by definition S holds iff she occurrently and reflectively believes both Fa and ~Fa. The Referentialist might adopt some *deflationary strategy* in order to make this *prima facie* unacceptable consequence somehow more palatable. There are two ways this might go.

The first strategy is to motivate a denial of (C) by reiterating the story about pragmatic implicatures. Consider what Kripke [1994, p. 368] thought about (C):

"We may suppose that [Lois Lane]...is a leading philosopher and logician. [She] would never let contradictory beliefs pass. And surely anyone, leading logician or no, is in principle in a position to notice and correct contradictory beliefs if he has them. Precisely for this reason, we regard individuals who contradict themselves as subject to greater censure than those who merely have false beliefs. But it is clear that [Lois Lane]...is in no position to see, by logic alone, that at least one of [her] beliefs must be false. [She] lacks information, not logical acumen. [She] cannot be convicted of inconsistency: to do so is incorrect."

Kripke is surely right about the latter: there is no way for Lois Lane to figure out that she holds logically contradictory De Dicto beliefs before the relevant empirical information is in. But for that very reason he is wrong about the former: the logical properties of her beliefs cannot be recovered just by reflection - even if, as we assume, her cognitive abilities were idealised. And the Implicature Theory helps us to see why. Although it is true that Lois Lane believes that Clark Kent does not fly and that Clark Kent flies, she assents only to 'Clark Kent does not fly'. She assents to 'Superman flies', but can consistently dissent from 'Clark Kent flies', because she dissents from 'Superman is Clark Kent'. She would have been "subject to greater censure" had she assented to 'Superman is Clark Kent', but the fact that she does not suffices to make perfectly good sense of her theoretical rationality.¹⁷ Why then does Kripke think that at least idealised speakers have reflective access to the logical properties of their De Dicto beliefs? Well, he must think that the contents of such beliefs are *Epistemically Transparent* in a way that allows for speakers to come to know such properties in an altogether non-empirical manner. But it is clear that speakers cannot hold logically contradictory De Dicto beliefs and not violate norms of rationality if their contents are fully Transparent to them. This means that if the Referentialist chooses to discard (C) such that speakers can rationally hold contradictory De Dicto beliefs, then the contents of such beliefs must be *Epistemically Opaque*.¹⁸ Indeed, this should follow in any case. We have seen that, on this view, true De Dicto belief ascriptions are semantically equivalent to their structural De Re counterparts. But since De Re beliefs are clearly Epistemically Opaque - S can consistently believe of a that it is both F and not F - then so

¹⁷ Note, however, that on the Implicature Theory, 'Lois Lane believes that Superman is Clark Kent' must be true, since it follows via (PS*) from the truth of 'Lois Lane believes that Superman is Superman' - contrary to what the legend tells us!

¹⁸ Epistemic Opacity was the claim that S can have introspective knowledge that she believes that P and that she believes that Q, and yet not have introspective knowledge of basic logical properties of her beliefs such as whether P is identical to, or distinct from, Q; cf. Sec. 1.3.

should the corresponding De Dicto beliefs also be.¹⁹ Lois Lane is unaware that she reflectively and occurrently believes a singular proposition and its negation in that she cannot introspectively discriminate between them due to the different ways in which she takes them. The upshot is that on the Referentialist account, the propositional contents of beliefs cannot be subject to Epistemic Transparency. In Sec. 1.3 I left it open whether Epistemic Transparency was an indispensable ingredient in Privileged Access. On the one hand, there is the worry that if we are ignorant of such key inferential relations between the contents of our beliefs as identity or contradiction, then there is a robust sense in which our intuitions about how those contents are transparently available to us are compromised. On the other hand, there is the thought that if indeed the contents of our beliefs are singular, as Referentialism tells us, then the corollary that we lack introspective access to logical properties of those contents should be welcome. After all, they are still Transparent in the weaker sense that in most cases when we have beliefs, we can have introspective knowledge of them.

3.4. Kripke's Paradox

This may prompt the Referentialist to pursue the second deflationary strategy: the problem about inconsistent beliefs just levelled against Referentialism is a deep problem that afflicts everyone. To appreciate the difficulty, Kripke [1994, p. 375] invites us to imagine the following: a competent speaker Peter learns the name 'Paderewski' with an identification of the person named as a famous pianist. Having learned this, Peter assents to 'Paderewski had musical talent', and we infer that Peter believes that Paderewski had musical talent. Later, in a different context, Peter learns of someone called 'Paderewski' who was a Polish Prime Minister. Peter, who unawares has just acquired two co-referring names, thinks that no

¹⁹ In short, the content of a belief is *Epistemically Opaque* iff the terms that occur in the embedded sentence that expresses that content are *referentially transparent*; see also Boghossian [1994].

politician has musical abilities, and so assents to 'Paderewski had no musical talent', and we infer that Peter believes that Paderewski had no musical talent. We can regiment this - *Kripke's Paradox* - as follows:

- | | |
|--|------|
| (i*) S is fully rational | ass. |
| (ii*) 'a' in context G and 'a' in context H are co-referring | ass. |
| (iii*) S assents to 'a is F' and S assents to 'a is not F' | ass. |
| (iv*) S believes that a is F and S believes that a is not F | (D) |
| (v*) S has contradictory beliefs, so S is not rational | (C) |

Both of (C) and (D) are *prima facie* warranted yet jointly inconsistent. But if the conjunction of (C) and (D) is false, then they cannot be used to refute the application of (PS*) in the Strengthened Intentional Argument; one cannot run a *reductio* that rests on false premises. In other words, why blame (PS*) in the Strengthened Intentional Argument, when we have *independent* reason to think that either or both of (C) and (D) are false? But if, as we suppose, (C) and (D) are platitudes that everyone should endorse, then problems about inconsistent beliefs should arise in any case - regardless of one's take on (PS*).

On closer examination, however, it emerges that Kripke's Paradox does not succeed in defusing the threat posed against Referentialism by the Strengthened Intentional Argument. Ironically enough, that Paradox makes the difficulties seem even more insuperable. Here is why. What is it to hold inconsistent beliefs? We agreed that it is not to hold contradictory De Re beliefs since the logical properties of such beliefs are not Epistemically Transparent. It is feasible to believe of a both that it is F and that it is not F without being reproached with rational inconsistency. Nor is there necessarily any inconsistency involved in holding De Dicto beliefs with embedded contradictory sentences. If the two occurrences of 'a' in (iv*) were not co-referring, then there would be no reason to think S held inconsistent beliefs. S can without inconsistency simultaneously believe that the bank is closed and that the bank is not closed, as long as what S believes is that the money-bank is closed and that the river-bank is open. What does seem rationally

inconsistent is to reflectively and occurrently believe the *propositions* expressed by 'a is F' and 'a is not F', where 'a' is unambiguous. Thus in the Strengthened Intentional Argument S holds contradictory beliefs as she believes the singular propositions expressed by those sentences. Nevertheless, S infringes no norms of rationality: although S assents to 'a is F', she does not assent to 'a is not F', and so does not realise that the propositions she believes are contradictory. Is a similar response available in the case of Kripke's Paradox? It seems not. S assents to both 'a is F' and 'a is not F'. Neither of S's beliefs in (iv*) carry any false implicatures and yet they have straightforward contradictory propositional contents. Faced with these difficulties, the only way to maintain Referentialism is therefore to resort to Epistemically Opaque contents, i.e. to give up on introspective knowledge of basic logical properties of occurrent mental content. And the worry is that core parts of what we intuitively take Privileged Access to be are thereby jettisoned.

Is the problem general, and, in particular, is Descriptivism faced with a similar problem? Well, let me briefly sketch how the logical form of De Dicto belief ascriptions is cashed out according to Descriptivism. If we assume that (RPB) is endorsed, then S believes that P states a relation of belief between S and the descriptive proposition expressed by 'P'. On this view, the contents of propositional attitudes are descriptive propositions consisting of descriptive, conceptual elements which are *modes of presentation* of the referents, if any, of the embedded terms.²⁰ Consequently, an ascription like (10) amounts to something like:

(19) B(Lois Lane, <m_S, m_F>)

²⁰ There is a very difficult issue about what a mode of presentation is that I eschewed in Chap. 2. I confess I have no detailed theory at hand. What I propose is that 'mode of presentation' be functionally defined along what Schiffer [1992, pp. 502-3] has called *Frege's Constraint*: (i) a rational person S may believe and disbelieve that a is F only if there are distinct modes of presentation m and m' such that S believes a to be F under m and disbelieves a to be F under m'; (ii) there are distinct modes of presentation m and m' such that S believes a to be F under m and disbelieves a to be F under m' only if S fails to realise that m and m' are both modes of presentation of a.

i.e. a two-place relation that obtains between Lois Lane and the proposition expressed by 'Superman flies' consisting of a particular mode of presentation of Superman and a particular mode of presentation of the property of flying. Since it is true that Superman is presented to Lois Lane in Superman-mode as the disguised hero who flies around and chases criminals, (19), i.e. (10), is true. (11), however, is false. Lois Lane does not believe the descriptive proposition expressed by 'Clark Kent flies' since Superman is not presented to her in Clark Kent-mode - m_{CK} - as the shy journalist who tries to chat her up. What this means is that (PS*) does not license replacement of 'Superman' for 'Clark Kent' *salva veritate*. So, the Descriptivist denies the validity of the application of (PS*) in line (v) in the Strengthened Intentional Argument.

According to Descriptivism, a referring term like 'Superman' is simply short for a cluster of descriptions which express identifying properties: 'the guy with the superman-properties'. It is then obvious that (PS*) fails to validate substitution of co-referring terms in true De Dicto belief ascriptions, since it is uncontroversial that definite descriptions are generally not subject to intersubstitution in belief contexts *salva veritate*. In order to apply (PS*) in true De Dicto belief ascriptions, one must ensure that the referents of the two terms in question are presented to the believer under the same mode of presentation. Conversely, Descriptivism does not deny intersubstitution of co-referring terms in true De Re belief ascriptions. Two De Re belief ascriptions are equivalent as long as the objects that individuate them are identical - no matter how they are referred to.²¹ What is claimed is that belief contexts do exhibit a genuine semantic De Re - De Dicto distinction. The truth of a De Dicto belief ascription does not license the inference of the corresponding De Re belief ascription. This follows from the fact that *existential generalisation* fails on such an ascription, that, say, the truth of (12) is not entailed by the truth of (10). If no one uniquely

²¹ So, one might say, there is an extent to which Referentialism accepts that belief contexts are referentially opaque, namely when definite descriptions occur in the embedded 'that'-clauses, and there is an extent to which Descriptivism accepts that belief contexts are referentially transparent, namely when those contexts are De Re.

has the Superman-properties, then it is possible for Lois Lane to believe that the guy with the Superman-properties flies without it being true that she believes of anyone in particular that he flies. If, on the other hand, a De Re belief ascription is true, then the truth of the corresponding De Dicto belief ascription cannot invariably be inferred. It follows that *some* De Dicto belief ascription will be true, since one cannot have a belief of an object without having some way of thinking of it, but it need not be the structural equivalent of the particular De Re belief ascription.²² It is true that Lois Lane believes of Clark Kent that he flies, but not that she believes that Clark Kent flies. The truth of the former is grounded in her true belief that Superman flies. The ascription 'Lois Lane believes that Superman flies' correctly reports the mode in which Lois Lane is presented with Superman when it comes to her belief that he flies - a belief which issues in her assent to 'Superman flies'.

In the case of Kripke's Paradox, a similar analysis explains away the appearance of paradox. The fact that (PS*) is not to blame in Kripke's Paradox as it does not feature there as a premise, does not mean that it is blameless in the Strengthened Intentional Argument. The dialectic is not that Kripke's Paradox shows that the conjunction of (D) and (C) is false *simpliciter*, hence also false in the Strengthened Intentional Argument, hence that (C) and (D) cannot be used as premises in a *reductio* of (PS*).²³ Unlike the Referentialist, the Descriptivist can endorse (C). (C) says that if S is fully rational, then S does not believe the proposition that a is F and the proposition that a is not F. On the Descriptivist account, (C)

²² As Dummett [1978, pp. 124-8, 1991, pp. 126-32] has argued, there can be no such thing as *bare knowledge of reference* of an expression: one cannot know of a particular object *o* that a name 'a' refers to it unless, for some F, one has an antecedent capacity to identify *o* as the F and knows that a is the F. One's predicative knowledge of *o* that 'a' refers to it rests on one's propositional knowledge that a is the F. See also Jackson [1998b, p. 216] and Stalnaker [1999b, p. 547].

²³ Kripke himself interpreted the Paradox in this way and since he also thought both (C) and (D) were indisputable, he had "...no firm belief as to how to solve it"; [1994, p. 369]. Many have, in my view incorrectly, followed Kripke in thinking that since Kripke's Paradox does not rest on (PS*), it is not what is at fault in the Strengthened Intentional Argument; for recent statements see Sosa [1996], Francis [1998] and compare with Kripke [1994, pp. 376-9].

simply does not apply in Kripke's Paradox, since there is no descriptive proposition associated with 'a is F' whose negation is associated with 'a is not F'. In (iv*), the content of S's belief that a is F consists of modes of presentation of a and F-ness and likewise for the content of S's belief that a is not F. But it follows from the way the Paradox was set up that S is presented with a in different modes. Peter, for instance, thinks of Paderewski in the politician-way and in the musician-way. So, S does not hold logically contradictory beliefs, so it does not follow by contraposition on (C) that S is less than fully rational.²⁴ On the Referentialist account, however, S does hold logically contradictory beliefs since the singular proposition expressed by 'a is not F' is the negation of the singular proposition expressed by 'a is F'. So, if S is not to be deemed irrational, then the Referentialist is compelled to concede that mental content is Epistemically Opaque. In other words, (C) must have counterexamples.

In the Strengthened Intentional Argument, however, (C) applies on the Descriptivist account since nothing indicates that 'a is F' and 'a is not F' in (v) should be associated with different descriptive propositions. Lois Lane, for instance, is presented with Superman in

²⁴ Sosa [1996] has proposed a response to Kripke's Paradox in much the same spirit. His diagnosis is that 'Paderewski' is ambiguous in a way the Referentialist cannot see. When can an ordinary language name be represented as a logical constant ϕ such that $[F\phi]$ and $[\sim F\phi]$ are logically contradictory? Not when ϕ is ambiguous. When is ϕ ambiguous? Well, certainly when ϕ has more than a single referent. Sameness of reference is a necessary condition on unambiguity, e.g. money-bank and river-bank. But, pace the Referentialist, it is not a sufficient condition. In order for an expression to occur unambiguously, it must also have the same semantic content, and, on the Descriptivist account, the two occurrences of 'a' in (iv*) have different descriptive contents. So, S does not have contradictory beliefs in (iv*). In [1994, p. 371], Kripke anticipates this response and argues in the case of Pierre's beliefs that *Londres* is pretty and that London is not pretty "...that the puzzle can arise even if Pierre associates to '*Londres*' and 'London' *exactly* the same *uniquely identifying* properties." Thus if Pierre associates the unique property of being the largest city in England, then he may just think that expressing his beliefs in English and French, 'England' and '*Angleterre*' name distinct countries. I think Kripke is wrong. Pierre does not associate *exactly* the same identifying properties: he associates with '*Londres*' the property of being the city called '*Londres*' by his French speech-community, but he does not associate that property with 'London'; and vice versa. In the case of 'Paderewski', it is even more difficult to see how to restate the paradox.

different modes, but they are modes which have distinct names. She does not associate different descriptive contents with tokens of 'Superman' that she falsely takes to pick out different individuals. So, were it not for the failure of (PS*), S would have been deemed irrational by (C). The Referentialist, however, is in no position to blame (PS*), and so must deny (C): the singular proposition expressed by 'a is not F' is the negation of the singular proposition expressed by 'a is F', but both can rationally be believed if mental content is Epistemically Opaque.

The Referentialist may retort that we have not answered the question: does S, or does S not, believe that a is F? Maybe the Descriptivist has a semantics for belief reports on offer which dissolves the Paradox when it comes to S's belief, but a Paradox remains when it comes to *our description* of S's belief. The point should be granted. We simply do not have adequate linguistic means to answer the question. It is like asking someone 'did you go to the bank?' in a context in which it is unclear whether the speaker means a money-bank or a river-bank. There is no determinate answer until after an interpretation has clarified the speaker's intended meaning. One must augment one's linguistic resources and say 'yes, I went to the money-bank'. By the same token, we could ascribe to S the belief that a, i.e. the G, is F, and the belief that a, i.e. the H, is not F. That way, we attribute true, consistent beliefs and remain faithful to S's linguistic behaviour.²⁵

3.5. The Strengthened Kripke's Paradox

What about our disquotational principle (D)? Well, if we could set up a similar argument against Referentialism that did not even rest on (C), then presumably it would be no help pointing out that fully rational speakers can hold contradictory beliefs if only their contents

²⁵ The objection is, of course, Kripke's [1994, p. 369]. For discussion on this point see Forbes [1990, pp. 558-63].

are Epistemically Opaque. What we need is to bear in mind that we can read the bi-conditional in:

(D) A competent speaker S is, on reflection, sincerely disposed to assent to 'P' iff S believes that P,

from left-to-right - from disposition to assent to belief - as we have done so far, but also from right-to-left - from belief to disposition to assent. It is important to see that (D) allows for cases in which speakers are *agnostic* as to whether they believe that P or \sim P. If S neither assents to, nor dissents from, 'P', then S is agnostic about P, and her refusal to assent or dissent is merely an indication of her demurral with respect to the belief that P. S's refusal to assent to 'P' does not imply that S believes that \sim P as she might refuse to assent to ' \sim P' as well. And her refusal to dissent from 'P' does not imply that she believes that P as she might refuse to assent to 'P' as well. Would S's dissent from 'P' suffice to entail her belief that \sim P? If we suppose that dissent is simply refusal to assent, then the answer is no. If S dissents from 'P', then S disbelieves that P, but that is not to say that S believes that \sim P, since S may also dissent from ' \sim P'. (D) respects agnosticism in that all contraposition on the right-to-left reading entails is that if S is not disposed to assent to 'P', then, *ceteris paribus*, it is not the case that S believes that P. (D) does not allow entailment from lack of - disposition to - assent to 'P' to the belief that \sim P. With this in mind, it would seem that we can use (D) and the story about Paderewski, to get a *Strengthened Kripke's Paradox*:

- | | |
|--|----------|
| (i**) S assents to 'a is F' | ass. |
| (ii**) S believes that a is F | (D) |
| (iii**) \sim (S assents to 'a is F') | ass. |
| (iv**) \sim (S believes that a is F) | (D) |
| (v**) Λ | \sim E |

When Peter learns of Paderewski as the famous pianist, he assents to 'Paderewski had musical talent', and we infer that he believes that Paderewski had musical talent. But when

he later learns of Paderewski as the Prime Minister, he refuses to assent to, or perhaps he dissents from, 'Paderewski had musical talent', and we infer that he lacks the belief that Paderewski had musical talent. So, given this story, the assumptions are harmless, and (D) seems innocent. But in line (v**) we have a contradiction, so something must give way. What can that be?

On the Referentialist account, the two occurrences of 'a is F' in (ii**) and (iv**) express the same singular proposition since the two occurrences of 'a' are co-referring. So, the application of the rule $\sim E$ (infer Λ from $P \ \& \ \sim P$) should be valid as we agreed that logical contradictions are between propositions and their negations. And, as I mentioned, no story about Epistemic Opacity is available to explain away the appearance of paradox: (ii**) and (iv**) will be contradictory regardless of the epistemic status of the propositions expressed by 'a is F' in (ii**) and (iv**). In Kripke's Paradox, the contradiction was between the propositions expressed by the embedded sentences 'a is F' and 'a is not F', but in the Strengthened Kripke's Paradox, the contradiction is between the propositions expressed by the belief reporting sentences 'S believes that a is F' and ' $\sim(S$ believes that a is F)'. So, in the latter, contradiction will arise no matter how Epistemically Opaque the proposition expressed by 'a is F' is.²⁶

I think there are two responses. First of all, the Referentialist may disavow (D) at least in its bi-conditional form such that refusal to assent, or dissent, does not indicate lack of belief. Lois Lane, for instance, would refuse to assent to, or perhaps even dissent from, 'Clark Kent flies', but according to Referentialism, she nevertheless believes that Clark Kent flies. However, the worry is that if there were no systematic links, even in the best cases, between assertoric use of sentences and expression of belief, then our practice of relying on our fellow speakers' linguistic behaviour as guiding our interpretation of their beliefs would seem to be unwarranted. Secondly, it may be objected that (D) is irrelevant to the Paradox since in fact (i**) and (iii**) are already contradictory. Suppose that on day 1

²⁶ It is instructive to note that Kripke [1994, p. 361] himself endorsed (D).

S assents to 'It's raining today', but on day 2 S refuses to assent to 'It's raining today'. Can we then use $\sim E$ to derive contradiction from: S assents to 'It's raining today' and $\sim(S$ assents to 'It's raining today')? The claim is that $\sim E$ fails to apply since somehow the time of utterance must be taken into account. Had S, in the very same context, both assented to, and refused to assent to, the very same sentence, then $\sim E$ would have applied. The point should be taken. Context-relativity clearly helps to solve our Paradox: S believes that a is F in context G and $\sim(S$ believes that a is F in context H) are not contradictory. But the worry is that S's beliefs are not sensitive to contexts. The mere fact that S lacks the belief that it is raining today does not entail that S ceases to believe that it was raining yesterday. Similarly, Peter's belief that Paderewski had musical talent is not relative to the context in which he learns of the famous pianist. When he later acquires the belief that Paderewski had no musical talent, he retains his old belief since he is unaware that his new belief is about the same individual.²⁷

In any case, the Strengthened Kripke's Paradox is easily accounted for if Descriptivism is endorsed. The rule $\sim E$ fails to apply since 'a is F' in (ii**) and 'a is F' in (iv**) express distinct descriptive propositions due to the fact that S associates distinct descriptive contents with the two occurrences of 'a'. The proposition that the famous pianist had musical talent does not contradict the proposition that the Polish Prime Minister had no

²⁷ Moore [1999] thinks (D) breaks down in cases where the context of assent is relevantly different from the context of attribution. Belief reports should therefore be contextualised: 'Peter believes that Paderewski had musical talent' is true relative to context G. But why should Peter's assent to 'Paderewski had no musical talent' in context H change the truth-value of the ascription in context G when Peter takes the information acquired in context H as *irrelevant* to his belief that Paderewski had musical talent? After all, Peter is unaware that 'Paderewski' in the two contexts are co-referring. Moreover, in context H where Peter learns of the Polish Prime Minister, he would still assent to 'Paderewski had musical talent' if asked about the famous pianist. Peter would thus be saddled with contradictory beliefs in the very same context.

musical talent. The upshot is that Referentialism is faced with difficulties that do not arise for Descriptivism.

Let me take stock. I have presented arguments against Referentialism to the effect that referring terms cannot be purely Referential, but must have associated descriptive content, and I have rejected a three-pronged defence of Referentialism. First I presented the Intentional Argument against Referentialism. Widespread intuitions have it that often 'S believes that a is F' is true while 'S believes that b is F' is false even though a is b, but how can that be if 'a' and 'b' are purely Referential? In response, the Referentialist invoked the Implicature Theory to the effect that our intuitions pertained to pragmatics and not semantics. Once proper account is taken of the defeasible pragmatic implicatures that arise when cognitive attitudes are ascribed, the universal applicability of (PS*) which follows from Referentialism, poses no problem. We answered by reinforcing the Intentional Argument. (PS*) in conjunction with *prima facie* platitudes about consistency and disquotation allow us to ascribe to fully rational speakers logically contradictory beliefs. The Referentialist's second response was then to acknowledge that (PS*) did issue in contradictory beliefs, but somehow deflate the implications for psychological rationality. The reason fully rational speakers can hold contradictory *De Dicto* beliefs is that their contents are Epistemically Opaque to them. Moreover, it need not be a devastating consequence that we must surrender the Epistemic Transparency of mental content. The third response was to concede that Referentialism sanctioned the existence of psychologically inconsistent beliefs, but to maintain that so did Descriptivism. Kripke's Paradox shows that we can deduce inconsistent beliefs from our platitudes without using (PS*). But we have not seen reason to suspect that the problem is general. According to Descriptivism, referring terms occur in opaque position in intentional contexts. This means that (PS*) disallows universal substitution of co-referring terms in such contexts, and so the Strengthened Intentional Argument is blocked. Also, the contents of beliefs are descriptive propositions consisting of modes of presentation. But Kripke's Paradox contains no

contradiction between a descriptive proposition and its negation, and so poses no problem on the Descriptivist account. Finally, we tried to strengthen Kripke's Paradox. If we could deduce a contradiction between the belief reporting sentences, and not just between the embedded clauses, then it would be no help pointing out that mental content is Epistemically Opaque. The Referentialist is again faced with difficulties that do not arise for the Descriptivist.

This suggests that we better settle for Descriptivism when it comes to a viable account of belief ascriptions. Referentialism is committed to the claim that co-referring singular terms are intersubstitutable in all contexts without change in truth-value, but we have seen strong grounds for thinking that this claim is false when it comes to intentional contexts, and that Descriptivism can explain why. That is, there are good reasons for suspecting that Referentialism is wrong in allowing us to use (PS*) to substitute all co-referring singular terms in intentional contexts *salva veritate*. However, unlike (PS), (PS*) was supposed to be a principle without counterexamples, so something has gone wrong. Correctly put, Descriptivism does not reject (PS*), only the Referentialist interpretation of the clause "...if the reference of α in $E(\alpha)$ is the same as the reference of β in $E(\beta)$." According to that interpretation, the reference that α and β have in E , for all E , is the reference that α and β have when not in E . In other words, sameness of reference in opaque contexts is sameness of reference in transparent contexts. Reference is not context-dependent.²⁸ This means that the clause in (PS) "...if α and β have the same reference" would do just as well. So, what the Descriptivist claims is that (PS) may fail across all intentional contexts, but that (PS*) holds provided the contextual constraint is given an adequate reading. According to - at least one version of - Descriptivism, sameness of reference in opaque contexts is sameness of propositional, i.e. descriptive, content in transparent contexts. This is what is meant by Frege's dictum [1994b] that terms occurring within the embedded sentence in an attitude attribution take *indirect reference* to their

²⁸ We are still ignoring quotational contexts.

customary descriptive contents (sense). It is easy to see how this move enables us to solve the Strengthened Intentional Argument: 'Superman' and 'Clark Kent' are not replaceable, because they differ in descriptive content, hence are not co-referring when the context E contains a 'believe that' operator. The substitution-instances in (10)/(11) and (iv)/(v) fail to satisfy the conditions for successful application of (PS*) which is to say that 'Superman'/'Clark Kent' and 'a'/'b' occur in opaque positions.

It thus seems as if we should endorse the logical form of belief ascriptions - S believes that a is F - that Descriptivism proposes, namely a two-place relation between a believer S and a descriptive proposition expressed by the embedded content-clause 'a is F':

(20) $B(S, \langle m_a; m_F \rangle)$.

Note that, on this account, (*Compositionality of Content*) is left intact: the propositional content of 'S believes that a is F' is uniquely determined by the propositional contents of its constituent expressions and their mode of combination. In particular, if 'a' and 'b' are (ordinarily) co-referring, but associated with distinct contents, then substituting 'a' for 'b' in 'S believes that a is F' yields a different propositional content.

The worry with this proposal is, however, that not all descriptive contents as ways of latching on to things in thought are intersubjectively shared amongst speakers. If in (10) Lois Lane associates a descriptive content with 'Superman' that is not part of common knowledge - if she thinks of Superman in a highly idiosyncratic way - then no one can believe or even understand what she believes when she believes the proposition expressed by 'Superman flies'. And moreover, as in the case of 'Paderewski', the particular modes of presentation which the believer associates with the embedded singular terms often cannot simply be read off from the speaker's ascription due to the insufficiency of the mere wording. Yet, as we shall see in Section 3.6, there is frequently no need to know *which* mode of presentation is employed to ensure successful communication, as long as *some* mode of presentation is shared. To require mutual knowledge of all associated descriptive

contents - in the face of community-wide variation - for successful communication is to run an unnecessary risk of communal misunderstandings in language.²⁹

To remedy this drawback, it may be suggested that descriptive contents be *quantified* over rather than *referred to*.³⁰ This allows us to maintain *Semantic Innocence* (SI): utterances of embedded sentences in belief attributions express just the propositions they would if not embedded. Typically, the plausibility of (SI) resides in considerations to do with theoretical economy, but in this context, it enables us to respect the modal rigidity of the embedded singular terms: they cannot be used to speak of the descriptive contents associated with them, but only to refer to their bearers. That is, a referring term exhibits the same semantic behaviour whether it occurs inside or outside the scope of an intentional operator. On this account, the logical form of (10) amounts to something like:

(22) $(\exists m_S) (\exists m_F) (m_S \text{ is a mode of presentation of Superman} \ \& \ m_F \text{ is a mode of presentation of fly-hood} \ \& \ B(\text{Lois Lane}, \langle m_S, m_F \rangle))$,

where modes of presentation are existentially quantified over. One of the advantages of this view is that one can know the truth-conditions of a De Dicto belief ascription like (10) without knowing the particular descriptive content - the mode of presentation - that the believer associates with the embedded singular term.

I shall not spend time developing this view in any detail. Instead I will finish this Chapter by giving an argument why I do not think Descriptivism jeopardises communication. This argument is important for another reason as well. What we have shown so far, if the arguments are good, is that singular terms in *opaque* position must have truth-conditionally relevant descriptive content. The semantic function of 'a' in 'S believes that a is F' cannot be merely to pick out an object. But we have not shown that singular

²⁹ There are also worries about the logical form of intuitively valid arguments like (a) Bill believes that Jones is tall; (b) Jones is tall; so (c) Bill believes something true. If, as on this view, 'Jones' in (a) refers to the descriptive content of 'Jones', while 'Jones' in (b) refers to Jones, then the argument equivocates and so is invalid.

³⁰ Cf. Forbes [1990], Noonan [1979] and Carruthers [1983].

terms occurring in *transparent* position must also have truth-conditionally relevant descriptive content. Maybe the semantic function of 'a' in 'S believes of a that it is F' is just to pick out an object. In other words, one could hold a *Hybrid View* according to which Referentialism is true in all but intentional contexts. All an utterance of an atomic sentence can express is a singular proposition, but an utterance of a true De Dicto belief attributing sentence states a three-place relation $B(S, P, m)$ between a speaker S, a singular proposition P, and a *semantically relevant* mode of presentation m of P.³¹ What I shall try to argue is that modes of presentation are semantic properties of singular terms regardless of the linguistic context in which they occur, and that they must be intersubjectively shared by speakers. So, we should not settle for less than full blown Descriptivism.

3.6. Communication

In the last Section I mentioned a worry that any version of Descriptivism must face: if a speaker's S understanding of a statement involving a referring term N requires her to be in possession of some way of thinking of the referent of N, then how can the informational content of that statement be passed on to another speaker S* if S* employs some other and very different way of thinking of the referent of N? In particular, if those 'ways of thinking' are not standardly reflected in any mutually shared meaning of N, then what reason does S have for thinking that by N she refers to the same thing as S* does when he uses N? All the Intentional Argument shows, if sound, is that S and S* must associate some descriptive

³¹ What is wrong with Referentialism is thus not that it dismisses modes of presentation altogether, but that it robs them of their semantic significance. It must be part of the semantic content of sentences in that-clauses to specify the way S takes the proposition P. Descriptivism, on the other hand, is right that associated descriptive content, understood as modes of presentation, enters into the truth-conditions of true De Dicto belief attributions, but wrong in thinking that nothing else does. The Hybrid View has over the years been advocated in various guises, see for instance McGinn [1999], Richard [1983], Recanati [1997], Crimmins & Perry [1989] and Schiffer [1992, 1995].

content with N, but if this content is not intersubjectively shared, and they have no other effective means by appeal to which they could resolve their differences, then, for all they could tell, they might be talking about different things, and so might fail to transmit any information by means of sentences containing N. It would thus seem, with Dummett's words [1978, p. 130], cf. [1991, p. 133], that the Argument has:

"[...] a major defect: it has no tendency to show that the sense of a word is a feature of the *language*. It shows, at best, that each speaker, if he is to associate a reference with a word, must attach a particular sense to it; it does not show any necessity for different speakers to attach the same sense to any one word, so long as the senses which they all attach to it determine the same reference. It therefore leaves open the possibility that the sense of a word is not part of its meaning at all, if meaning is to be something objective and shared by all speakers [...]"

In other words, what the Intentional Argument shows is that two co-referring singular terms N and M must have associated different descriptive contents *when* they occur in opaque contexts. It is open to hold, as indeed most proponents of the Hybrid View do [cf. fn. 31], that when N and M occur in transparent contexts, as in 'N is F' and 'M is F' where 'Fx' is extensional, N and M merely express the same singular content. But, the thought continues, [1978, p. 132], cf. [1991, p. 133]:

" [...] the use of language for communication [...] depends upon the informational content of a sentence being constant from speaker to speaker. If language is to serve as a medium of *communication*, it is not sufficient that a sentence should in fact be true under the interpretation placed on it by one speaker just in case it is true under that placed on it by another; it is also necessary that both speakers should be aware of the fact."

If Dummett is right, what is called for is some additional argument that N and M must have associated different descriptive contents regardless of the linguistic context in which they appear, and moreover that such contents are "constant from speaker to speaker." What must

be argued is thus that descriptive content is needed not just in a *theory of speaker's reference*, but also in a *theory of reference for a common language*. If we could show not only that every speaker must associate some descriptive content with the terms they use for singular reference, but also that they must associate approximately the same descriptive content with those terms, then we would have reason to think that descriptive content is "a feature of the language". Consider, for a start, Frege's example³²:

Suppose...that Herbert Garner knows that Dr. Gustav Lauben was born on 6 September 1875 in N.N. and [that] this is not true of anyone else; suppose, however, that he does not know where Dr. Lauben now lives nor indeed anything else about him. On the other hand, suppose that Leo Peter does not know that Dr. Lauben was born on 6 September 1875 in N.N. Then as far as the proper name "Dr. Gustav Lauben" is concerned, Herbert Garner and Leo Peter do not speak the same language, although they do in fact refer to the same man with this name; for they do not know that they are doing so.

Let us call the problem that Frege raises the *Problem about Communication*³³: how must the descriptive contents - as ways of thinking of objects - which different speakers associate with given names, be related if communication is to be possible? Let me illustrate. Suppose Garner truly says: 'Dr. Lauben is wounded', in the presence of Peter. Peter assumes that by 'Dr. Lauben' Garner refers to the same man Peter uses 'Dr. Lauben' to refer to. Peter is right and so comes to believe that Dr. Lauben is wounded. Is this a successful case of communication? Well, if Referentialism is right about the semantics of transparently occurring names, then the sole propositional content of the sentence is the singular proposition <Dr. Lauben; wounded-ness> which Peter apprehends upon hearing Garner's utterance. Since what Garner said is what Peter understood, it would seem that information has successfully been transferred from Garner to Peter. Referentialism is thus committed to the view that preservation of reference suffices for understanding if by 'understanding' we

³² Cf. [1994a, p. 523].

³³ Cf. also Heck's [1994, pp. 81, 92, 99].

mean grasp of propositional content. As long as Peter gets the reference right, he also gets the propositional content right, and so, he will understand what Garner has said, despite the fact that they have no identifying knowledge in common. Consequently, Peter will come to believe what Garner believes if he has no reason to think Garner is wrong. According to Referentialism, the answer to our question is therefore that:

(A) We need only get the reference right: mutual reference is necessary and sufficient for mutual understanding, hence for communication.³⁴

The worry with (A) is not the necessity-claim: preservation of reference is imperative if we are to rely on others' testimony as a means of acquiring knowledge. It is rather that mutual reference does not suffice for securing *transmission of knowledge* by means of language. It is true, as the example shows, that as long as Garner and Peter refer to the same man, they run no risk of distorting information, i.e. of not *transmitting true beliefs*. The problem is, as Dummett and Frege emphasise, that although Garner and Peter refer to the same man by 'Dr. Lauben', they do so unwittingly. Moreover, they have no effective means by which they could come to know of their co-referential use of 'Dr. Lauben', since there is no overlap in their identifying knowledge of Dr. Lauben. But we tend to think that communication requires understanding and that understanding is a species of knowledge such that to understand what someone has said is to know what he has said. *The purpose of communication is therefore to facilitate transmission of knowledge and not just true beliefs.*³⁵ In our example, Peter does not know what Garner has said, because he does not know who 'Dr. Lauben' refers to in Garner's utterance, even though 'Dr. Lauben' *accidentally* refers to the same man in Peter's idiolect. Garner's statement prompts Peter to believe what Garner said but incidentally so; for all Peter knows, Garner might have been

³⁴ Kripke [1994, pp. 359-60] seems to have advocated this view.

³⁵ In general, if S knows that P, and S asserts 'P' in the presence of S*, then, if S* understands the assertion and accepts it as true, then, *ceteris paribus*, S comes to *know* that P; cf. Heck [1994, pp. 91-4], [1996, pp. 155-56] and Evans [1982, pp. 310-11].

speaking of someone else. Peter's belief that Garner has said that Dr. Lauben is wounded is true, but *true belief is not knowledge*. Peter has no justification to believe that he believes what Garner believes. Nor is there any reliable method that he could appeal to. Peter cannot rely on the wording: it is a coincidence that Garner and himself use orthographically identical names of Dr. Lauben. There is no intrinsic connection in Peter's and Garner's use of 'Dr. Lauben' apart from the fact that Dr. Lauben is causally responsible for their respective use of that name. Also, 'Dr. Lauben' may be unusual, but most types of ordinary proper names, e.g. 'Anna' or 'John', have multiple referents. So, it seems that all Peter is entitled to infer from Garner's statement is that, if true, someone named 'Dr. Lauben' has been wounded.³⁶ But he cannot disquote since nothing he can point to indicates that he has understood Garner's use of it. So, although Peter gets the reference right, he does not fully grasp its propositional content, so that content cannot be exhausted by its reference. Or to make the point more obvious, suppose Peter knows of Dr. Lauben only under his other name 'Gustav Hendricks'. Imagine also, as in the Superman story, that Peter has no reason to think that 'Dr. Lauben' and 'Gustav Hendricks' are co-referring. As it turns out, Peter forms the belief that Gustav Hendricks is wounded, upon hearing Garner's assertion of 'Dr. Lauben has been wounded'. According to (A), communication achieves its purpose in securing that Peter acquires Garner's true belief. But surely, it does not. Peter cannot come to know that Gustav Hendricks is wounded on accepting 'Dr. Lauben is wounded' as true when he has no evidence that Gustav Hendricks is Dr. Lauben. This suggests that communication can misfire even if reference is preserved. Although Peter gets the reference right, he has not understood Garner's statement, and so, if understanding is

³⁶ This is where the other-dependent descriptions enters the picture; cf. Sec. 2.2. All Peter understands by Garner's use of 'Dr. Lauben' is 'the individual whom Garner is intending to refer to, on this occasion, with his use of 'Dr. Lauben'. Such 'buck-passing' descriptions are especially relevant when one acquires the mastery of names not hitherto encountered. They show the possibility of competent use of names without possessing any uniquely identifying information other than pointing to the causal chain by which that use is passed on.

knowledge of propositional content, there must be more to the content of a name than its referent. So, perhaps (A) should be replaced by:

(B) We need only know that we have got the reference right: knowledge of mutual reference is necessary and sufficient for mutual understanding, hence for communication.

Thus if Peter knows that by 'Dr. Lauben' Garner refers to the man Peter uses 'Dr. Lauben' to refer to, then Peter understands Garner's statement since he knows that Garner says of Dr. Lauben that he is wounded. So, if Garner's statement is true, and Peter has no reason not to accept it as such, then, *ceteris paribus*, Peter comes to know that Dr. Lauben is wounded. Now, how does Peter come to know that they are talking about the same man? I said above that Peter could not rely on the mere wording yet that seems to be what we do all the time. I truly say: 'Bill Clinton is on ITV' and you accept my assertion and come to know that Bill Clinton is on ITV. In normal contexts, you assume I mean the US president, and not, say, your uncle who goes under the same name. In the normal run of things, this common identifying knowledge is *contextually presupposed* and never actually brought out. But it could be. If you had just told me about your uncle, it might be unclear whom I was now talking about. For each of the Clintons, we would no doubt have different identifying knowledge, but there will also be enough common ground between us to resolve the issue. Suppose instead I truly say: 'John was late' in a context in which we both know that I could have meant either of two men. In order for you to come to know what I said, you must know which one I intended, i.e. disambiguate my utterance. Communication succeeds only after I have drawn on common associated properties: 'I meant John, the German, not John, the Englishman'. So, you must think of the referent of my use of 'John' in such a way that you can come to know whom I was referring to, and it is hard to see how you could do that if we were not thinking of that referent in relevantly similar ways; if you did not also think of John as, say, the German. Had you by chance formed the belief that John (the German) was late upon hearing my utterance without bothering about whom I intended to pick out,

your true belief would fall short of knowledge. So, there are cases in which understanding requires not only that speakers think of objects in similar ways but also *know* that they do. Hence, the way to knowingly get the reference right is for there to be common identifying knowledge that is accessible if needed. (B) should therefore be modified by:

(C) We need to share *some* identifying knowledge: mutual descriptive content is both necessary and sufficient for mutual understanding, hence for communication.³⁷

Consider two putative counterexamples to (C). First, the *necessity-claim*³⁸: Suppose Lingers understands with Peter by the name 'Dr. Lauben' the only doctor who lives in a house known to both of them, and suppose furthermore that Lingers understands with Garner by 'Dr. Lauben' the unique individual born 6 September 1875. Then one might think that both Garner and Peter could communicate with Lingers, although not with each other. But this is false: if Lingers reports Garner's utterance to Peter, then Peter will come to know, on the basis of a third-party testimony, that Garner uses 'Dr. Lauben' to refer to the same man that he uses 'Dr. Lauben' to refer to, and so Peter could come to know what Garner expressed by his utterance. There are two things to say in response. First, were it not for the intermediary, Peter and Garner could not successfully have exchanged information, but we are interested in the basic cases where two speakers purport to communicate. In such cases, it is still true that knowledge of mutual reference goes via knowledge of mutually associated descriptive content. Peter, for instance, is able to communicate with Lingers only because they share identifying knowledge of Dr. Lauben. Secondly, it seems to me they could share some identifying information by way of deferring to Lingers, i.e. with 'Dr. Lauben' they could both associate 'the man whom Lingers uses 'Dr. Lauben' to

³⁷ (C) assumes that enough descriptive content is in common between speakers to guarantee uniqueness. E.g. you and I can both associate the description 'the current President' with 'Bill Clinton' and yet disagree as to which nation has Bill Clinton as their current President. In order to enable transmission of knowledge between us by means of that name, we would need to associate something like 'the current US President'.

³⁸ This example is modelled on Carruthers [1983, pp. 26-7] and Frege [1994a].

refer to'. Then suppose Garner knows from Lingers that Dr. Lauben is wounded. Upon hearing Garner's utterance of 'Dr. Lauben is wounded', Peter could thus come to know that whoever Lingers uses 'Dr. Lauben' to refer to is wounded.

Secondly, the *sufficiency-claim*³⁹: Suppose I saw a very drunk American in the Keys last night and decided to call him 'Jack'. After I left, you saw the same guy, also dubbed him 'Jack' and associated the same identifying information with the name. Later on we met up and you told me what happened in the Keys: 'Jack got thrown out', which prompted me to believe that Jack got thrown out. Despite shared descriptive content, I have not understood you, so my belief, albeit true, does not amount to knowledge. I think the example shows that when the context provides insufficient information for the speakers to know that they are talking about the same thing, then, just as in the example with 'John', they must *know* that they share the same identifying information. Suppose I had named him 'Jones', and suppose your utterance of 'Jack got thrown out' somehow caused me to believe that Jones got thrown out. Maybe I only saw one drunk in the Keys last night and therefore assumed you were talking about him. It seems clear that although my belief that Jones got thrown out is in fact true, it does not constitute knowledge since I do not know that the individual you refer to by 'Jack' is the individual I refer to by 'Jones'. But the only way to know that is to know that by 'Jack' and 'Jones' we associate the same identifying information: the drunk American in the Keys last night. So, it appears that knowledge of shared descriptive content would suffice to secure successful communication after all.

The point is thus that if we are to transmit knowledge by means of singular terms, then there must be some constraints on, or uniformity in, the kind of descriptive contents that we associate with them.⁴⁰ In the case of general terms like 'water', 'lemon' or 'sofa', one can expect a community-wide consensus in associated descriptive content, and in the case

³⁹ This example is modelled on Byrne & Thau [1996, p. 147]. Heck [1994, pp. 101-2; 1996, pp. 153-4] accepts (C) as far as the necessity-claim is concerned but is uncertain about the sufficiency-claim.

⁴⁰ Cf. also Heck [1994, pp. 101-4].

of proper names a less extensive and more speaker-relative agreement. Communication cannot take place between speakers who think about objects in radically different ways. On the contrary, successful communication about objects requires intersubjectively shared ways of thinking about them. But if speakers by and large associate the same descriptive contents with the same singular terms, then, as Dummett says, there is reason to think that descriptive contents are properties of such terms as used in a common language, and not merely what each individual speaker must attach to terms if she is to use them to refer to objects. And if descriptive contents are features of singular terms as such, then they will be so regardless of the linguistic context in which those terms are embedded. Whether a singular term occurs in transparent or opaque position, it should have descriptive content associated with it.

Let me sum up. I have argued that Referentialism is wrong in claiming that the only propositional content a singular referring term has is its reference. Descriptive content as modes of presentation of objects must be invoked to account for the truth-conditions of sentences ascribing propositional attitudes. Moreover, I have tried to show how considerations about the enabling conditions for successful communication support the stronger claim that descriptive content also plays a role when referring terms occur in *transparent* position. Regardless of the linguistic context in which a term occurs, it must have widespread associated descriptive conditions by satisfaction of which an object is referred to by that term. So, Descriptivism seems indeed to be a viable option. I have argued against the Modal Argument against Descriptivism in Chapter 2, and in favour of the Intentional Argument against Referentialism in this Chapter. What I shall now do is present a version of Descriptivism - *Causal Rigidified Descriptivism* - which not only accounts for the Modal Argument but also for the Twin Earth Argument.

Chapter 4. Two-Dimensionalism

4.1. Causal Rigidified Descriptivism

I have presented arguments against Referentialism to the effect that the propositional content of a referring term *N* cannot consist merely in its having the reference it has. It must also have associated descriptive content which goes into the truth-conditions of sentences containing *N* in opaque as well as in transparent position. Descriptive content must be invoked to account for the truth-conditions of all sentences and not only those which ascribe propositional attitudes. The question we should now address is: what is descriptive content? We started off by assuming that the descriptive content of *N* was simply given by a set of definite descriptions that competent speakers associated with *N*. These definite descriptions were supposed to encapsulate the way in which those speakers were thinking about the referent of *N*, i.e. the mode of presentation of that referent. Nothing was initially presupposed about the character of those descriptions, except that if *N* had a referent, then it would be whatever object uniquely satisfied those descriptions. The descriptive content of, say, 'water' was given by a cluster of stereotypical properties that all and only typical samples of the natural kind water possess - what we have called 'the watery stuff'. But it is plain that our conception of what goes into the descriptive content of

'water' would have to take into account the arguments brought forward against Descriptivism. Consider this formulation of the *Twin Earth Argument*:

- (1) Suppose Twin Earth is a remote place in W_A just like Earth except that XYZ is the watery stuff.
- (2) Suppose 'water' is short for 'the watery stuff'.
- (3) Then my utterance on Earth of 'water is wet' is true at Twin Earth iff the watery stuff on Twin Earth is wet.
- (4) But my utterance on Earth of 'water is wet' is true at Twin Earth iff H_2O is wet.
- (5) So, 'water' is not short for 'the watery stuff'.

On the assumption in (2) that 'water' is a purely Descriptive term, 'water' picks out all and only watery stuff. In particular, my use of 'water' on Earth picks out the watery stuff XYZ on Twin Earth. So, my utterance of 'water is wet' is true at Twin Earth due to the fact that XYZ is wet on Twin Earth. But our intuition is that water is not wet on Twin Earth since indeed there is no water there. Although XYZ is watery, XYZ is not water, and so the fact that XYZ is wet cannot make my utterance of 'water is wet' true; only facts about H_2O can make such utterances true. Obviously, (1)-(5) is not much of an argument on its own as (4) merely contradicts (3), but our intuitions seem to support (4): if water is H_2O on Earth, then water is H_2O everywhere, including Twin Earth.¹ Moreover, the truth of (4) is clearly unacceptable by Descriptivist lights in that it gives singular truth-conditions. According to Descriptivism, one need not know the referent of 'water' in order to understand the truth-conditions of sentences containing 'water'. There is, however, a way of accommodating our intuitions without buying into Referentialism. It is at this stage worthwhile noting that Putnam [1996, p. 19] took his own thought-experiment to prove that:

¹ Moreover, as I argued in Sec. 1.2, 2.3, 2.4, such intuitions are firmly grounded in the scientific role that natural kind terms play in so-called theoretical reductions. When physical science discovered that water is H_2O , it discovered which micro-physical property is causally responsible for the macro-physical properties that water have, e.g. why it obeys certain physical laws.

"words like "water" have an unnoticed indexical component: "water" is stuff that bears a certain similarity relation to the water *around here*."

I think the remark is spot on. The conditions under which my token of 'water is wet' is true on Earth, or on Twin Earth, must concern the kind of watery stuff I am *acquainted* with on Earth. It is integral to the concept of water that water is the stuff I drink, swim in, look at, listen to, and what not, i.e. that water is the watery stuff of my acquaintance. Given that Twin Earth is taken to be a remote place in W_A , I, qua Earthling, have had no causal encounters with XYZ on Twin Earth, and so by 'water' I do not refer to XYZ even if XYZ is watery. The only watery stuff Earthly speakers sustain causal connections with is H_2O , and so H_2O is what they refer to by 'water'. Similarly, my doppelgänger, who is acquainted with the watery XYZ on Twin Earth, will by tokens of 'water' refer to XYZ, and not to H_2O even if watery.

What the Twin Earth Argument thus tells us is that *causality* must play a role in determining the reference of a term. In most ordinary cases, we use referring terms to pick out objects that not only have all the superficial properties we are familiar with, but also are the objects we are causally related to. Subjectively indistinguishable counterparts in far off locations are rarely the kind of objects we intend to pick out. The point is that the Descriptivist should allow for a causality constraint to be built into the purely descriptive content: 'water' refers to the watery stuff of our acquaintance. Correspondingly, sentences containing 'water' should be assigned causally constrained descriptive truth-conditions. In particular, (3) should read:

(3*) Then my utterance on Earth of 'water is wet' is true at Twin Earth iff the watery stuff of our acquaintance is wet.

Note how this gets the truth-conditions right with respect to Twin Earth. My token of 'water is wet' is true at Twin Earth iff the watery stuff on Earth is wet on Twin Earth. The watery stuff on Earth is H_2O and H_2O is not wet on Twin Earth since *ex hypothesi* there is no H_2O on Twin Earth. So, (3*) respects our intuitions about Twin Earth. Yet the truth-condition in

(3*) is not singular since the reference of 'water' does not appear on the right-hand side of the bi-conditional. What Twin Earth teaches us is thus not that 'water' is a purely Referential term, i.e. that the reference of 'water' cannot go via associated properties. It is true that Twin Earth tells us that 'water' cannot be purely Descriptive, but there is scope for intermediary views. What we learn is that causal properties play a role in mediating the reference of 'water', and that in odd cases they overrule the purely descriptive properties. So, we learn from Twin Earth which properties form the cluster associated with 'water', and we learn something about how to weigh them against each other in possible cases.

The same is true of 'Aristotle'. Suppose Aristotle had an epistemic counterpart on a distinct Twin Earth in W_A , and suppose I uttered 'Aristotle wrote *De Anima*'. Then, if 'Aristotle' were short for 'the famous philosopher of antiquity, etc.', then my utterance would be true at Twin Earth iff the counterpart wrote *De Anima* on Twin Earth. But I was clearly making an assertion about Aristotle. So, we must supplement the cluster of properties commonly associated with 'Aristotle' with a causality constraint: the individual causally responsible for my use of 'Aristotle'. I have had no causal encounters with Aristotle, but I trust that at the end of the causal chain leading to my use of the name, there is someone - Aristotle - who has all the properties I associate with the name. At some point Aristotle was baptised 'Aristotle' and that act set off a causal chain in which the name was passed on from speaker to speaker. Being competent with the name, I can appeal to this chain if need be. Here is an example. In my Department, there are three professors of moral philosophy called 'John'. Suppose I say: 'John attended the seminar'. You were not there, so you ask: 'which John?'. I reply: 'the famous philosopher'. You ask: 'which famous philosopher?'. I reply: 'the professor of moral philosophy'. You ask: 'which professor of moral philosophy?'. I reply: 'the author of *Weighing Goods*'. You get it: 'Oh, him!'. When speakers try to figure out what a given term is used to refer to, they resort to shared associated properties. In that pursuit, they are guided by the conviction that a unique individual is responsible for the particular use which that term has and which they both take

part in by associating sufficiently similar properties with that term in order to determine that they are referring to the same object. In cases where no unique referent is immediately identified, speakers can track down an object, as it were, by going back along the causal chain until they have identified a property that uniquely marks out that object, e.g. 'the individual born at location x at time t '. In rare cases, like water and twin-water, one would have to go back to the very dubbing itself before a referent could be determined. The only difference between 'water' and 'twin-water' is that distinct natural kinds are causally responsible for the way they are used, and so all a speaker can do by way of laying down a referent for his use of 'water' may be to appeal to the very causal chain that has led to this use.

Suppose now that Twin Earth is, not a remote planet in W_A , but a non-actual W_P . Twin Earth is not a real place far off in W_A , but a counterfactual W_P . Suppose also that we now know that the reference of 'water' must be causally constrained, i.e. that water is the watery stuff of our acquaintance. On these two assumptions, we can run the Twin Earth Argument once again:

- (6) Suppose Twin Earth is a counterfactual W_P just like Earth in W_A except that XYZ is the watery stuff.
- (7) Suppose 'water' is short for 'the watery stuff of our acquaintance'.
- (8) Then my utterance on Earth of 'water is wet' is true at Twin Earth iff the watery stuff of our acquaintance on Twin Earth is wet.
- (9) But my utterance on Earth of 'water is wet' is true at Twin Earth iff H_2O is wet.
- (10) So, 'water' is not short for 'the watery stuff of our acquaintance'.

It is important to see that causality does nothing to ensure we get the reference right with respect to counterfactual circumstances.² If Twin Earth is a remote place in W_A , then I and

² Putnam [1996, p. 19] failed to distinguish between these two ways of thinking of Twin Earth. His intention was to take Twin Earth as a "...Far-away planet in the actual universe..."; [1990, p. 60, cf. p. 69]. But he thought that the Twin Earth Argument showed that water is "...the water *around here*" where he took the expression 'around here' to function as a rigidifier; op. cit. It is true that the indexical 'here' is modally rigid,

my fellow Earthlings will not have had causal encounters with the watery stuff on Twin Earth, and so could not by our tokens of 'water' refer to XYZ. Only the inhabitants on Twin Earth have encountered XYZ, so only their tokens of 'water' will pick out that stuff. But if Twin Earth is taken as a counterfactual W_P , then *ex hypothesi* we causally interact with the watery XYZ on Twin Earth, and so, on the assumption that 'water' is short for 'the watery stuff of our acquaintance', our tokens of 'water' refer to XYZ on Twin Earth. We are asked to imagine a W_P in which *we* find ourselves implanted in a physical environment just like in W_A apart from the fact that it contains no water; only the superficially indistinguishable XYZ. So, in those counterfactual circumstances, we causally interact with XYZ yet we agreed our Earthly tokens of 'water' do not pick out XYZ even in such circumstances.

The Referentialist will now argue that since only the singular truth-conditions in (9) can account for such counterfactual Twin Earth cases, 'water' is a purely Referential term. The semantic function of 'water' is not to describe its referent, but merely to serve as a rigid designator of a natural kind. In light of our discussion in Chapter 2, the Descriptivist will, however, not be persuaded by such an argument. There is no inference from modal rigidity to singular content. All the Descriptivist needs to account for counterfactuals is rigidification. So, in order to safeguard against Twin Earth thought of as another W_P , the Descriptivist must build into the descriptive content of 'water', not only a causality constraint, but also a rigidification device: the *actual* watery stuff of our acquaintance. This gets the truth-conditions right with respect to counterfactual circumstances:

(8*) Then my utterance on Earth of 'water is wet' is true at Twin Earth iff the actual watery stuff of our acquaintance is wet.

Given that I am an Earthling, my tokens of 'water' will rigidly pick out H_2O since H_2O is the watery stuff of my acquaintance in W_A . So, the condition under which my utterance is true at a counterfactual Twin Earth is that H_2O is wet on this Twin Earth. The point is again

but the way he set up the Twin Earth Argument does not license any conclusion about the rigidity of 'water'. Burge [1979; 1982] was explicit about this difference.

that there is a way of accommodating our intuition about possible cases without buying into singular content. The truth-condition in (8*) is not singular since the referent of 'water' does not appear on the right-hand side of the bi-conditional. All we need is an appropriately constrained descriptive condition: a causality constraint when Twin Earth is taken in the remote-place sense, and a rigidity constraint when Twin Earth is taken in the another- W_P sense.

The argument (6)-(10) is of course nothing but the Modal Argument in a different guise. What we learn from that Argument is not that the reference of a term N cannot go via associated properties. On the contrary, we learn that what matters when N picks out an object o is that o has those associated properties in W_A . N refers to o iff o uniquely has the associated properties in W_A ; regardless of whether o has them in various W_P . What a counterfactual Twin Earth emphasises is that the correct description of a W_P often depends on facts about how W_A is, and not just on how that W_P is stipulated to be. Our intuition that 'water' is a rigid designator is based on intuitions about the correct description of Twin Earth: although XYZ is watery, it is not correctly described by us as 'water' given that the watery stuff on Earth is H_2O .

The same is true of 'Aristotle'. Suppose Aristotle has a doppelgänger in W_1 who has all the properties we standardly associate with 'Aristotle' in W_A - even the property of being the causal origin of our use of 'Aristotle'. Under which conditions is my utterance of 'Aristotle wrote *De Anima*' true at W_1 ? Clearly not iff Aristotle's doppelgänger wrote *De Anima* in W_1 . We use 'Aristotle' to speak of the same individual in all counterfactual circumstances. So, my utterance is true at W_1 iff the famous philosopher of antiquity, etc., in W_A wrote *De Anima* in W_1 . We learn from the Modal Argument that our description of W_1 depends on W_A : 'Aristotle' picks out Aristotle's doppelgänger at W_1 iff Aristotle's doppelgänger has all and only the properties we associate with 'Aristotle' in W_A . But given that he has not - Aristotle has all and only those properties in W_A - 'Aristotle' does not refer to Aristotle's doppelgänger at W_1 , no matter who has those properties in W_1 .

In short, neither the Twin Earth Argument nor the Modal Argument forces us to abandon descriptive content. On the contrary, we learn something about the character of the descriptions which encapsulate that content. The former tells us that at least some of those descriptions must be couched in a causal vocabulary, and the second tells us that those descriptions must be rigidified. What constitutes the cluster of associated properties are thus not only purely general, qualitative properties, but also causal properties and properties which objects have in W_A . We can call this view *Causal Rigidified Descriptivism*.³ I shall now spend some time on how 'actually' works.

4.2. The Logic of 'Actually'.

There are various ways a modally flexible expression can be rigidified. Take 'the teacher of Alexander', and suppose that Aristotle taught Alexander in W_A , but that Plato taught him in W_1 . The definite description will then pick out Aristotle at W_A and Plato at W_1 . By rigidification we ensure that the description picks out the same individual at all W_P . More precisely, that the individual who satisfies the description at W_A , is the individual who is being picked out at all W_P . A simple way of doing that is to qualify the description: 'the teacher of Alexander in W_A '. This accommodates what I have called:

(W_P -Rigidity) N is a rigid designator iff N refers to the same object o at every W_P in which o exists and not to something else at W_P in which o does not exist.

Note that we could, in principle, have rigidified the description by fixing on, not the actual, but a possible satisfier of the description. 'The teacher of Alexander in W_1 ' is modally rigid by (W_P -Rigidity): it picks out Plato at all W_P including W_A . Nevertheless, when it comes

³ This view does not have many friends in the literature; exceptions are White [1982], Lewis [1999a], Stanley [1999a] and Jackson [1998b].

to singular referring terms, proper names or natural kind terms, we wish to pick out the referents of these terms at W_A at all W_P . In the normal run of things, it is their actual referents that concern us when we consider what would happen had various circumstances obtained. So, it will suffice for our purposes to consider only rigidifying devices that fix on objects picked out at W_A . Instead of adding 'in W_A ' we could equally well have inserted an actuality-operator. The expression 'the *actual* teacher of Alexander' does also accommodate (W_P -Rigidity).⁴

When we use 'actually' for the purpose of rigidification, we are interested in a distinctive logical use of the expression in natural language - not in a mere pragmatic use as when the BBC-commentator says 'actually it was Shearer, not Owen, who got sent off...', where 'actually' merely serves to dispel confusion. One such logical use is to clarify matters of relative scope involving modal operators. Take Aristotle's:

(11) It is possible that a man who is sitting down should be standing up,

which is ambiguous between the false:

(12) $\Diamond \exists x (Fx \ \& \ Gx)$,

where F is 'sitting down' and G is 'standing up', and the true:

(13) $\exists x (Fx \ \& \ \Diamond Gx)$.

Another way to effect the consistent reading is to introduce 'actually'. Thus:

(14) It is possible that a man who is actually sitting down should be standing up,

disambiguates (11) by indicating that (13) is the more appropriate. There are, however, logical uses of 'actually' which seem to suggest that a genuine actuality-operator must be introduced into formal modal languages alongside \Diamond - and \Box -operators. Consider:

⁴ A third way of accomplishing the same result is to use the demonstrative 'this world'. The expression 'the teacher of Alexander in this world' is also modally rigid by (W_P -Rigidity). There may be some difficulties in spelling out how the demonstration is supposed to work. If I say 'this man is drunk', then my audience will understand me to be referring to the contextually salient adult male, but it is not clear that I can demonstratively identify a possible world in the same manner. I shall henceforth only use an actuality-operator. For more on the technical differences see Fitch [1981].

(15) It is possible that everything which is actually square should be blue,
and the following three putative renditions:

(16) $\Diamond \forall x (Sx \rightarrow Bx)$,

(17) $\forall x \Diamond (Sx \rightarrow Bx)$,

(18) $\forall x (Sx \rightarrow \Diamond Bx)$,

with progressively narrower scope for the \Diamond -operator with respect to the universal quantifier. None, however, will do. (16) says that there is a W_P in which all square things are blue, but (15) does not require the existence of a W_P in which some things are both square and blue. (17) says that for each thing there is a W_P in which it is blue if square, and (18) says that for each thing which is actually square there is a W_P in which that thing is blue, but (15) requires the existence of a W_P in which all things square in W_A are blue in that W_P . So, it is essential that the \Diamond -operator takes large scope with respect to the universal quantifier, yet that in determining the truth-value of the embedded universally quantified sentence, we are allowed to *refer back* to W_A and not just to how things are in W_P . A regimentation which acknowledges a semantically significant actuality-operator @ will have at its disposal such a back-referring device. Thus:

(19) $\Diamond \forall x (@Sx \rightarrow Bx)$,

translates (15) satisfactorily since (19) is true only if there is a W_P in which all things which are square in W_A are blue in W_P .⁵ In short, sentences like (15) suggest that modal sentences containing 'actually' cannot adequately be expressed within standard first-order quantified modal logic.⁶

In saying that (19) is an adequate translation of (15), it is assumed that 'actually' is given an *indexical analysis* to the effect that 'actually' rigidly refers to the worldly context in which it is uttered. We can use Stalnaker's [1999a] *semantic matrices* to represent the way 'actually' works. These matrices shall prove important in the following, so it is crucial

⁵ Proviso regarding existence, cf. Davies [1981b, p. 221].

⁶ The examples in this section are based on Crossley & Humberstone [1977] and Davies [1981b].

to understand how these devices function. What one must bear in mind is that there are two different ways in which a term may apply at W_P . As glossed by Jackson [1998a, p. 48], one must distinguish between:

...what the term applies to under various hypotheses about which world is the actual world, or [...] what the term applies to under various counterfactual hypotheses.

Let me illustrate. Suppose Plato is the teacher of Alexander and the author of *De Anima* in W_1 , and suppose that Aristotle taught Alexander and wrote *De Anima* in W_2 . Then I utter:

(20) The actual teacher of Alexander wrote *De Anima*,

in W_1 . What I said is true at W_1 since the description 'the actual teacher of Alexander' picks out Plato and he wrote that book in W_1 . Is (20) true at W_2 ? Well, it depends on how we think of W_2 . If we *think of W_2 as counterfactual*, then the answer is negative. 'The actual teacher of Alexander' refers to Plato at W_2 due to *the rigidifier 'actual'*, but he did not write *De Anima* in W_2 . If, however, we *think of W_2 as actual*, then the answer is affirmative. To think of a W_P as actual is to think of W_P as a possible context of utterance or acquisition - a W_P in which we are speaking. In that case, the description refers to Aristotle due to *the indexicality of 'actual'*, and he did write *De Anima* in W_2 . That is, if I utter 'the actual teacher of Alexander' in W_2 , then we take W_2 to be actual, and so I refer to Aristotle since we assume that he taught Alexander in W_2 .

This is the basic mechanism. When one thinks of a W_P as *actual*, one considers the possibility that W_P turns out to be W_A . When one thinks of a W_P as *counterfactual*, one takes a given way W_A has turned out, a fixed W_A , and then considers W_P as a counterfactual W_P . Other philosophers have made roughly the same distinction. Kaplan [1989], for instance, distinguished between a possible *context of utterance* and a possible *circumstance of evaluation*. A *context* is a location where a sentence is uttered consisting of

a time, place, speaker and a W_P . A circumstance is an *index* consisting of contextual features that need not go together in any possible context.⁷ Consider the sentence:

(21) I feel elated,

and suppose that S feels elated in W_1 but not in W_2 , and that R feels elated in W_2 but not in W_1 . As Kaplan famously argued, 'I' is an indexical which rigidly refers to the speaker of the context of utterance. The linguistic meaning of 'I' - its *character* - is a rule from a context of utterance to the speaker of that context. So, the linguistic meaning determines that if S is the speaker, then S is the referent of 'I' in (21). The singular proposition expressed in W_1 $\langle S; \text{elated-ness} \rangle$ is thus true at W_1 . Is (21) also true at W_2 ? It depends on how we think of W_2 . If W_2 is thought of as a circumstance of evaluation, i.e. as counterfactual, then, given that S is the speaker in W_1 , (21) is false at W_2 . When we evaluate sentences at W_P considered as counterfactual, we consider the truth-value at W_P of the singular propositions expressed in the W_P we consider as actual. It is thus false that S feels elated at W_2 . If, however, W_2 is thought of as a possible context of utterance, i.e. as a candidate for W_A , then our evaluation depends on who is speaking in W_2 . If S is the speaker, then S expresses the singular proposition $\langle S; \text{elated-ness} \rangle$ which is false at W_2 . If R is the speaker, then R expresses the singular proposition $\langle R; \text{elated-ness} \rangle$ which is true at W_2 . So, the indexicality of 'I' is a relevant property when we consider W_P as actual in that its propositional content is sensitive to variations in contexts of utterance. The modal rigidity of 'I', on the other hand, is a relevant property when we consider W_P as counterfactual in that the reference of 'I' at such W_P is fixed by whoever is the speaker in the W_P we consider as actual.⁸

⁷ See also Lewis [1981] for more on the distinction between context and index.

⁸ Note that for modally flexible expressions, it is irrelevant for the truth-value of a sentence at W_P how W_P is taken. With our examples, 'the teacher of Alexander is a philosopher' is true at W_P iff the teacher of Alexander at W_P is a philosopher at W_P ; it does not matter whether W_P is taken as actual or counterfactual. Flexible designators apply to a counterfactual W_P independently of how W_A turns out.

There is a danger that natural language is liable to confuse us as to which of these two ways of conceiving of W_P we are engaging in. I have so far been talking about the truth-value of a sentence *at a* W_P . But it is clear that this is potentially ambiguous between *the possibility of a sentence expressing a truth* and *the possible truth of what a sentence expresses*. It is, as Kaplan [1989, p. 290] emphasised:

"...likely to confuse what the truth-value of the proposition *actually* expressed would have been under some possible circumstances with what the truth-value of the proposition that would have been expressed *would have been* under those circumstances"

I shall therefore adopt the following terminology inspired by Prior [1976]. When we think of a W_P as counterfactual, we say that a sentence is *true-of* W_P , and when we think of a W_P as actual, we say that a sentence is *true-in* W_P . If a sentence is true-in a W_P , then it expresses a truth when uttered in that W_P , but a sentence can be true-of a W_P without occurring as uttered in it. If a sentence is true-of a W_P , then whatever it expresses when uttered in W_A is true of that W_P . A sentence can be true-in a W_P without being true-of the W_P in which it is true, and a sentence can be true-of a W_P without being true-in the W_P of which it is true. Let me illustrate. Take:

(20) The actual teacher of Alexander wrote *De Anima*,

but now suppose that Plato is the author of *De Anima* in both W_1 and W_2 . Plato is still the teacher of Alexander in W_1 and Aristotle is his teacher in W_2 . Now, (20) is true-in W_1 . When we consider W_1 as actual, then 'the actual teacher of Alexander' refers to Plato, and Plato wrote *De Anima* in W_1 . Is (20) also true-of W_1 ? Well, in order to find out whether a sentence is true-of a W_P one must first determine which W_P is considered as actual; one must, as it were, fix an W_A . If we let W_2 be actual, then 'the actual teacher of Alexander' refers to Aristotle. But since Aristotle did not write *De Anima* in W_1 , (20) is false-of W_1 .

So, (20) is true-in W_1 , but is, when W_2 is taken as actual, false-of W_1 .⁹ (20) is, however, false-in W_2 . If we consider W_2 as actual, then 'the actual teacher of Alexander' refers to Aristotle, but Aristotle did not write *De Anima* in W_2 . Is (20) also false-of W_2 ? Well, if we fix W_1 as actual, then 'the actual teacher of Alexander' refers to Plato, but Plato did write *De Anima* in W_2 . So, (20) is false-in W_2 , but is, when W_1 is taken as actual, true-of W_2 .¹⁰

The same is roughly true of:

(21) I feel elated.

Suppose S is the speaker in W_1 and R is the speaker in W_2 , and suppose that S feels elated in both W_1 and W_2 and that R feels unhappy in both W_1 and W_2 . Then (21) is true-in W_1 but false-in W_2 . And if we consider W_1 as our context of utterance, then (21) is true-of W_2 . In W_1 'I' refers to S and S does feel elated in W_2 . If, however, we take W_2 as our context of utterance, then (21) is false-of W_1 . In W_2 'I' refers to R but R does not feel elated in W_1 .

Note that when we consider the truth-value of a sentence as uttered in different possible contexts, we assume the sentence is given a *fixed* interpretation. It is not that 'I feel elated' means that the speaker of the context feels elated in W_1 and that pigs fly in W_2 . In

⁹ A natural question to ask is: can a sentence be true-of a W_p in the special case where the same W_p is fixed as actual? The answer is clearly yes. If W_1 is taken as actual, then 'the actual teacher of Alexander' refers to Plato and Plato wrote *De Anima* in W_1 . It follows trivially that if a sentence is true-in W_p , then it is true-of W_p provided W_p is taken as actual. Or as Lewis [1981, p. 88] puts the same point: a sentence s is true at context c iff s is true at c at the index of the context; the context of utterance and the context of evaluation, i.e. the index, are simply one and the same.

¹⁰ For another example of a sentence which is true-of a W_p , but is false-in that W_p , see Dummett [1981, pp. 567]. Note that Dummett's distinction between truth *with respect to* a world and truth *in* a world corresponds to my distinction between truth-of a world and truth-in a world. Or take Priori's own example [1976]: 'all propositions are negative'. This sentence is presumably true-of a W_p . But if it were to occur in that W_p it would not be possible that all propositions in that W_p are negative since it itself expresses an affirmative proposition.

this case, the linguistic meaning of 'I' is a fixed rule from a context of utterance to a speaker.¹¹

What I have said about truth also holds for reference. So far I have been talking about the reference of a term at a W_P but once we allow for the possibility that reference is sensitive to the context of utterance or acquisition, this is potentially misleading between what it refers to at counterfactual W_P and what it would refer to if used in W_P . For simplicity, I shall adopt the same terminology and say that an expression ψ refers to o of a W_P when W_P is considered as counterfactual and that ψ refers to an object o in a W_P when W_P is considered as actual. With our terminology in place, we can then say that @ is an operator which has the main semantic feature that for any referring expression ψ , @ ψ refers to o of a W_P just in case ψ refers to o in W_A . But given that we have allowed for the possibility that another W_P turns out to be actual, what @ ψ refers to depends on which W_P we consider as actual. E.g. 'the actual teacher of Alexander' refers to Aristotle of a counterfactual W_P just in case 'the actual teacher of Alexander' refers to Aristotle in the W_P we consider as actual. In general, the reference of an expression thus has a *double world-dependency*. It depends on the W_P we consider as counterfactual and on the W_P we consider as actual. In order to represent this kind of double world-dependency on reference, what is called for is not the single indexed function $F_1: W_P \rightarrow \text{Extension/Reference}$, from Sec. 2.2, but a doubly indexed function:

$$F_2: W_{PA} \times W_{PC} \rightarrow \text{Extension/Reference},$$

¹¹ What I do not have in mind is nicely brought out by the riddle: if you call a horse's tail a leg how many legs does a horse have? The answer is, of course, four! Yet had 'leg' applied to tails as well as to legs, the answer would have been five. In order to appreciate the other way of taking the question we must change the interpretation placed on 'leg', i.e. the way we use 'leg' in our speech community. No such changes are called for in the case of expressions containing 'I' or 'actual'. For a different example see Sec. 4.3.

where the first parameter represents W_P considered as actual¹² and the second parameter represents W_P considered as counterfactual. The truth of a sentence has a similar double world-dependency. The truth of 'I have a headache' depends on who the speaker is and whether he has a headache. If in W_A I speak and have a headache, then the sentence is actually true. By changing the context of utterance, while fixing the circumstance of evaluation, we can change the truth-value. If in W_A you do not have a headache, then the sentence would be false if uttered by you. Or we can change the truth-value by changing the circumstance of evaluation, while fixing the context of utterance. If in a W_P I do not have a headache, then the sentence would be false if uttered by me. In other words, the facts determine the truth-value of what was actually expressed by a sentence, but had the facts been different, something different might have been expressed and that, in turn, might have had a different truth-value.

We can now picture this double world-dependency by a *semantic matrix*. Along the horizontal axis are W_P i and j qua counterfactual, i.e. qua circumstance of evaluation, and along the vertical axis are W_P i and j qua actual, i.e. qua contexts of utterance or acquisition. Truth, on the matrix, is thus dependent on both ways of thinking of W_P . A *horizontal line* represents the propositional content of an expression in a given W_P and each item in the row records the reference of that expression of the W_P heading the vertical column in which it occurs. A *vertical line*, on the other hand, represents the variations in the reference of an expression effected by shifts in W_P qua contexts of utterance or acquisition. According to the indexical analysis of 'actually', utterances of the definite description:

(22) The actual world,

¹² Due to Putnam-type Twin Earth cases - where Twin Earth is a remote planet in W_A - we really need centred W_P in the argument place W_{PA} in F_2 . That is, since one and the same utterance type can occur in different contexts within the same W_P , W_{PA} better be a W_P with a designated spatio-temporal location, if it is to get the reference right in all contexts of acquisition. Alternatively, F_2 could take as argument a token utterance or, as I propose, include an acquaintance-condition; cf. also Chalmers [1996, p. 60].

can thus be depicted by the matrix in fig. 1:

fig. 1	i	j
i	i	i
j	j	j

Due to the indexicality of 'actual', the reference of 'the actual world' varies as the W_P qua actual varies, but once we have fixed a W_P qua actual, 'the actual world' refers to the same W_P at all counterfactual W_P . In our terminology, the first horizontal row shows that 'the actual world' refers to i in i and to i in i of j . The second horizontal row shows that 'the actual world' refers to j in j and to j in j of i . Likewise, utterances of (20) can be pictured by the matrix in fig. 2:

fig. 2	W_1	W_2
W_1	True	False
W_2	False	True

We now suppose that in W_1 Aristotle taught Alexander and wrote *De Anima*, and that in W_2 Plato taught Alexander and wrote *De Anima*. The first horizontal row shows that 'the actual teacher of Alexander wrote *De Anima*' is true-in W_1 , but is, in W_1 , false-of W_2 . In W_1 Aristotle taught Alexander and wrote *De Anima*, but he did not write it in W_2 . The second horizontal row shows that 'the actual teacher of Alexander wrote *De Anima*' is true-in W_2 , but is, in W_2 , false-of W_1 . In W_2 Plato taught Alexander and wrote *De Anima*, but he did not write it in W_1 .

(21) can also be represented by fig. 2 on the assumption that in W_1 S speaks and feels elated, and in W_2 R speaks and feels elated. The first horizontal row would then show that 'I feel elated' is true-in W_1 , but is, in W_1 , false-of W_2 . S is the speaker in W_1 , but S only feels elated in W_1 . The second horizontal row shows that 'I feel elated' is true-in W_2 , but is, in W_2 , false-of W_1 . R is the speaker in W_2 , but R only feels elated in W_2 .

Let me give another example. Assume, as we standardly do after Kaplan [1989], that the expression 'this' is an indexical which rigidly refers to the demonstratively identified place or entity in the context of utterance. Now consider the sentence:

(23) This is the actual world.

It is a consequence of the indexical analysis of 'actually' that (23) is true-of all W_P in all W_P . (23) thus expresses a necessary proposition, i.e. a proposition which is true at every W_P . This is depicted in fig. 3 by the fact that at least one horizontal line contains nothing but truths:

fig. 3	W_1	W_2
W_1	True	True
W_2	True	True

The same is true for sentences like 'If Aristotle exists, then Aristotle was the actual teacher of Alexander' or 'I am the actual speaker' which state identities between two rigid designators. In general, any indicative sentence s containing 'actually' will be, if true, then necessarily so. In other words, on the indexical analysis, $\forall s (@s \rightarrow \Box@s)$ will hold good.¹³

¹³ If we take S5 to be the background modal logic, then $@s \rightarrow \Box@s$ will be an axiom (A5) in S5@. Moreover, $\Box@s$ and $@s$ are necessarily equivalent, since we have that $\Box@s \rightarrow @s$ by a substitution-instance of the T-axiom of necessity $\Box s \rightarrow s$. All (A5) says is that if $@s$ is true, then it is true-of every W_P that $@s$ is true, but this is perfectly compatible with the possibility that $@s$ is false-in some other W_P . That is, had the sentence $@s$ been uttered in a different W_P , $@s$ might have expressed a falsehood. To couch this other notion of necessity, we need a fixedly-operator \mathcal{F} which combines with $@$ such that $\mathcal{F}@s$ is true iff whichever W_P plays the role as actual, s would have been true-in that W_P . Accordingly $\mathcal{F}@s$ and $\Box@s$ are equivalent, but only for $@$ -free formulae s , so that if the background logic for \mathcal{F} is S5@, $\mathcal{F}@s \leftrightarrow \Box@s$ (for any $@$ -free formula s) will be an axiom in S5@ \mathcal{F} . \mathcal{F} , however, does not by itself represent necessity: $s \rightarrow \mathcal{F}s$ is true, but only for all $@$ -free formulae s , since $@s \rightarrow \mathcal{F}@s$ is an invalid substitution instance. To be sure, \mathcal{F} was precisely introduced to invalidate $@s \rightarrow \mathcal{F}@s$, so that $\mathcal{F}@$ would form a notion of necessity distinct from \Box . What represents necessity is $\mathcal{F}@$: what is necessary is what would be true no matter which W_P is considered as actual. In short, $\Box@s$ says that s is true-of all W whereas $\mathcal{F}@s$ says that s is true-in all W_P . For more on this two-dimensional modal logic see Davies & Humberstone [1981a], Crosswell & Humberstone [1977], Segerberg [1973] and Stalnaker [1999a].

But it may be objected that this involves a kind of unacceptable *fatalism*.¹⁴ For surely, it is not necessary that the world we call 'actual' is W_A . W_A might have been different in various ways, that is, another W_P might have been W_A : $\forall W_P (W_P = W_A \rightarrow \Diamond W_P \neq W_A)$. Likewise, it is neither necessary that Aristotle taught Alexander in W_A nor necessary that I was the speaker of some sentence token in W_A . Plato might have been Alexander's teacher in W_A and you might have been the utterer of that token.

The objection is good, but so is the response. First of all, I have used ' W_A ' as short for 'the actual world'. This immediately invites the caveat not to confuse the true wide scope reading 'the actual world might not have been the actual world' with the false narrow scope reading 'it might have been that the actual world is not the actual world'. To say that ' \Box (23)' is true is just to say that it is not possible that this world, W , is not W_A , i.e. had W_A been different from the way it is, it would not have been W_A but another W_P . There is, however, also a sense in which (23) allows for the possibility envisaged: of this world, W_A , it is possible that it is not W_A , i.e. a different W_P might have been W_A . In other words, the proposition expressed by (23) in W_A is necessary. But had another W_P been actual, i.e. had (23) been uttered in a different W_P , the expressions 'this' and 'the actual world' would have referred to a different W_P , and so (23) would have stated a different truth from the one stated in W_A , although still a necessary one. Likewise with 'Aristotle was the actual teacher of Alexander'. If Aristotle was Alexander's teacher in W_A , then there is presumably a W_P in which Aristotle did not teach Alexander, but there is no W_P in which Aristotle did not teach Alexander in W_A . The proposition expressed by that sentence is therefore necessary. If, however, we suppose that 'Aristotle' is short for 'the actual teacher of Alexander', then had 'Aristotle was the actual teacher of Alexander' been uttered in a W_P in which, say, Plato

¹⁴ As van Inwagen [1980] seems to think.

taught Alexander, then that sentence would have stated a different truth from the one stated in W_A , although still a necessary one.¹⁵

4.3. Two-Dimensionalism

Consider now the following example originally due to Evans [1982, 1996b]. We know that some individual uniquely invented the zip fastener, but we do not know who it was. It would thus be convenient to introduce a name for this individual. Imagine the stipulation:

(S*) Let 'Julius' refer to whoever invented the zip fastener.

Evans' idea was now that, following Referentialism, (S*) *fixes the reference* of 'Julius' such that 'Julius' functions as modally rigid by (W_P -Rigidity): 'Julius' refers to the unique satisfier in W_A of 'the inventor of the zip' of every W_P in which that satisfier exists and not to something else of W_P in which it does not. But, following Descriptivism, (S*) *also confers descriptive content* on 'Julius'. What this means is not only that it is a priori

¹⁵ There is an interesting issue about the metaphysical ramifications of the semantics of 'actually'. Lewis [1973, pp. 85-6] endorsed the indexical theory of 'actually' as part of his defence of extreme realism about possible worlds. His view had three components: (i) W_P , i.e. ways things might have been, exist; (ii) W_P are just more things of the same kind as the W_A ; (iii) W_P are not reducible but respectable entities in their own right. Lewis suggested that the actuality of W_A is a world-relative attribute; our world has it relative to itself but so do all the others. In contrast, Stalnaker [1976] has argued that one can endorse the indexical theory as a semantical thesis without such metaphysical commitments; just as someone convinced of the indexical analysis of 'I' could be a solipsist. He proposed a moderate realism about possible worlds which accepts only (i) and (iii). There exist W_P in the sense of there existing, in their own right, ways our world might have been, without there existing W_P which are the ways our world might have been. Stalnaker's point was that a thing cannot be identical with a property of that thing, so, in particular, a thing, e.g. the cosmos, whatever that is, cannot be identical with any way that thing could have been, including the way that thing actually is. W_A has the property of being the way it is, but W_A is *not* the way it is. That property is instantiated, but the property of being a way W_A might have been, while existing in its own right, is not.

knowable that the reference of 'Julius' is fixed by 'the inventor of the zip fastener'. What it means is that all it takes to understand 'Julius' as it occurs in sentences is that 'Julius' refers to whoever invented the zip. It is not constitutive of mastery of 'Julius' that one knows of the referent of 'Julius' that 'Julius' refers to him.¹⁶ Consider now:

(24) Julius was the inventor of the zip,

and suppose that Reagan invented the zip in W_1 , Bush invented the zip in W_2 , and Clinton invented the zip in W_3 . Suppose moreover that (S*) is made in each of W_1 , W_2 , and W_3 . Then the matrix in fig. 4 depicts utterances of (24):

fig. 4	W_1	W_2	W_3	W_4
W_1	True	False	False	True
W_2	False	True	False	False
W_3	False	False	True	False
W_4	False	False	True	False

Following Stalnaker [1999a, pp. 9-16, 80-4], we can define three propositions on the matrix. For each W_P in the vertical axis, i.e. qua actual, we can define a *horizontal proposition* as a function from a W_P in the horizontal axis, i.e. qua counterfactual, into a truth-value. For instance, if we take W_1 as actual, then the horizontal proposition is the function that yields the value true given W_1 as argument, the value false given W_2 as

¹⁶ In [1980, p. 59] Kripke complains that the Descriptivist uses 'sense' ambiguously between giving the meaning and fixing the reference of an expression. If Evans' example is coherent, then (S*) performs both tasks! For a similar example, see Dummett's [1973, pp. 112-32] 'St. Anne' as short for 'the mother of the virgin Mary', and his [1981, pp. 562-70] 'Deutero-Isaiah' as short for 'the author of prophecy embodied in chapters 40 to 55 of the Book of Isaiah'. Dummett held that all singular terms had descriptive content by means of which they picked out their referents, but due to the problems discussed in Section 3.6 about widespread variation in associated descriptive content, he chose examples where the linguistic community possessed only a very limited amount of identifying information. In the case of 'Julius', the name does simply not have a use prior to (S). The point of the examples was merely to demonstrate that names *could have* descriptive content. As we saw, other difficulties arise when we consider less contrived, ordinary names like 'Aristotle' or 'Bill Clinton'.

argument, etc. In each W_P (24) is thus false-of some W_P which is to say that (24) expresses a contingent proposition in each W_P . This is recorded in fig. 4 by the fact that each horizontal proposition contains some falsehoods. A *diagonal proposition* is defined as a function from a set of W_P consisting of a W_P in the vertical axis and the same W_P in the horizontal axis into a truth-value.¹⁷ It is the proposition associated with a sentence which for any W_P is true-of a W_P iff it is true-in that W_P , i.e. the proposition that is true for any W_P iff the horizontal proposition expressed in that W_P is true at that W_P . If we take W_1/W_1 as arguments, then we get the value truth, and likewise for W_2/W_2 and W_3/W_3 . The significance of the fact that the diagonal proposition in fig. 4 contains nothing but truths is that (24) expresses a true proposition in every W_P qua actual. So, someone who grasps the descriptive content of 'Julius' - knows of (S^*) - will know that (24) could only be used to express truths. For each W_P in the horizontal axis, we can finally define a *vertical proposition* as a function from a W_P in the vertical axis into a truth-value. For instance, if we take W_1 as counterfactual, then the vertical proposition is the function that yields the value true given W_1 as argument, the value false given W_2 as argument, etc. This shows that the truth-value of (24) is sensitive to variations in W_P qua actual even if we hold W_P qua counterfactual constant: in W_1 (24) is true-of W_1 , in W_2 (24) is false-of W_1 , etc. The reason for this is that, due to the hidden 'actuality' operator in 'Julius', the proposition expressed by (24) is itself sensitive to variations in W_P qua actual. We could thus define different horizontal propositions for each W_P qua actual.¹⁸

¹⁷ Note that a sentence can be assigned more than one diagonal proposition given that there are multiple contexts of utterance or acquisition within a single W_P - Twin Earth can be taken to be in W_A . That is why we really need as arguments, not W_P , but spatio-temporal locations within W_P ; cf. fn. 12.

¹⁸ It is important to emphasise that for each sentence numerous diagonal, horizontal and vertical propositions can be defined on the matrix for that sentence. It is only for the sake of convenience that I have confined the matrices to contain a very limited number of worlds. Note also how these three propositions elucidate the distinction between one- and two-dimensional operators. \Box is a one-dimensional operator which takes horizontal propositions into horizontal propositions: to evaluate \Box one need only look at the values of P on

There is, as I briefly mentioned in Sec. 4.2, another way of changing the proposition expressed, hence the truth-value of a sentence, and it is important not to confuse them. One can change the proposition expressed by (24) in a W_P just by changing the interpretation of 'Julius'. Suppose we had introduced 'Julius' by a different reference-fixing stipulation:

(S**) Let 'Julius' refer to whoever got elected US President in 1992,

and that W_4 is a W_P just like W_1 , except that in W_4 (S**), and not (S*), is made. Then, as fig. 4 shows, (24) is not true-in W_4 , so the diagonal no longer contains just truths. In W_4 'Julius' refers to Clinton, but Reagan, and not Clinton, invented the zip in W_4 . Does this mean that (24) is not true-in every W_P ? No. It is true that 'Julius' might have been used differently from the way 'Julius' *ex hypothesi* happens to be used, just like we might have used 'T' to denote the hearer, but such possibilities are semantically irrelevant. They tell us nothing about the language as we actually use and understand it. So, when we represent an expression by the matrix, we take it as having its standard community-wide interpretation in all the W_P in which it is used.

It is worth dwelling on what Evans took the semantic function of 'Julius' to be. 'Julius' is supposed to be a *descriptive name*, i.e. a singular term used rigidly to refer to an individual except that it has associated descriptive content which (i) mediates its reference and (ii) constitutes grasping its propositional content. This invites the thought that 'Julius' is just short for a rigidified definite description which encodes that content: 'the actual inventor of the zip'.¹⁹ The plausibility of this identification derives from the fact that, given (i) and (ii), 'Julius' and 'the actual inventor of the zip' should be intersubstitutable *salva veritate* in both modal and intentional contexts. Evans [1996b, pp. 187-92] held that 'Julius'

the horizontal line. @ is a two-dimensional operator which take a whole matrix into another whole matrix: the matrix for @P is the one that projects the diagonal proposition for P onto the vertical. @ thus turns a contingent truth into a necessary truth and a contingent falsehood into a necessary falsehood.

¹⁹ We assume uniqueness. If more than one individual invented the zip in W_A , then we must supplement the description with an acquaintance condition just like we did with 'water'.

and 'the actual inventor of the zip' were not semantically equivalent since he was sceptical as to whether definite descriptions - rigidified or not - belonged to the class of referring expressions. But as Davies & Humberstone [1981a, pp. 11-13] pointed out, it is hard to see what semantic considerations could possibly distinguish them. Evans presumably thought that the truth-value of a sentence containing an ordinary proper name like 'Aristotle' was not sensitive to variations in W_p qua actual.²⁰ But if it is true, as Evans himself seemed to think, that the reference of 'Julius' is sensitive to such variations, then so is 'the actual inventor of the zip'; the two expressions will thus invariably receive the same representations on the matrices.²¹

Is it nevertheless feasible to hold that 'Julius' is a rigid designator with descriptive content given by 'the inventor of the zip' yet deny that 'Julius' simply abbreviates 'the actual inventor of the zip'? It would seem that only if this combination of views is coherent, could the Descriptivist maintain that definite descriptions are not genuinely referring expressions. Bear in mind the two ways of responding to the Modal Argument from Sec. 2.5 and 2.6 respectively. We have so far assumed that the Descriptivist concedes the Rigidity Principle (RP) and meets the Modal Argument by rigidification. But, as I argued, Descriptivism could equally well have blocked the Modal Argument by refusing to accept (RP) on which that Argument hangs. In that case, there would be no need for rigidification. 'Julius' and 'the inventor of the zip' will have the same descriptive content yet presumably differ in modal properties. It is one thing, however, to establish that 'Julius' has descriptive content, but

²⁰ They are, in Evans' [1996b] terminology, *deeply rigid* designators, where an expression is deeply rigid iff it refers to the same object in all W_p , i.e. at all W_p qua actual. To my knowledge, nobody has argued persuasively that there are deeply rigid designators in natural language nor indeed what would ensure that there could be such designators. However, see fn. 45.

²¹ At pp. 206-7, *op. cit.*, he writes: "...to hold that the states of affairs of *y's being the inventor of the zip and being F*, for example, could serve to make the sentence 'Julius is F' true, would appear to commit one to the view that, had *y* invented the zip and been *F*, the sentence would have been true. But is this not inconsistent with the fact that the sentence 'Julius is F' is not true with respect to that situation? I accept the counterfactual claim, but there is no inconsistency in doing so."

another thing to secure its modal rigidity in the absence of a rigidifier. The worry is how 'Julius' could be modally rigid if its reference is mediated by a modally flexible description: if the reference of 'Julius' at counterfactual W_P is determined by whoever invented the zip in W_P , then 'Julius' will not be a rigid designator. What is called for is some restriction on the descriptive content such that all it does by way of determining reference is that it fixes the reference of 'Julius' in W_A . It is simply not a property of descriptive content to determine reference at counterfactual W_P . But this is exactly what rigidification does. It ensures that 'Julius' picks out an individual in counterfactual circumstances only if that individual is the actual inventor of the zip; regardless of whether that individual invented the zip in those circumstances. Without the aid of devices like 'actually', it is hard to see how the Descriptivist can avail herself of modal rigidity. How, for instance, could she get the W_P -truth-conditions for sentences like (24) correct when her pledge to descriptive content rules out appeal to singular truth-conditions? On this background, I shall henceforth assume that any viable version of Descriptivism must embrace a form of rigidification strategy.

So, assume that a singular term like 'Julius' is simply short for 'the actual inventor of the zip'. Given that 'actually' is both a rigidifier and an indexical, it is clear that neither (DR) from Sec. 2.2 nor (RR) from Sec. 2.3. will do as a principle by which the reference of 'Julius' is governed.²² According to (DR), a singular term N is not modally rigid, and according to (RR) the reference of N is not sensitive to variations in which W_P we consider as actual. In Sec. 2.2 and 2.3 I took it for granted that ' W_A ' was a *fixed* world - the *actual* W_A as it were. But Two-Dimensionalism allows for the possibility of considering different W_P as actual, and so N may refer to different objects as we vary W_P qua actual. In W_1 where Reagan invented the zip, 'Julius' refers to Reagan, but in W_2 where Bush invented

²² To repeat, Descriptivist Reference (DR) reads: N refers to o at W_P iff o satisfies ϕ at W_P , whereas Referentialist Reference (RR) reads: N refers to o at W_P iff o satisfies ϕ at W_A ; where N is a singular term, o an object and ϕ some descriptive conditions.

the zip, 'Julius' refers to Bush. So, when it comes to reference at counterfactual W_P , *Two-Dimensionalism* rules that the reference of N at counterfactual W_P is determined by satisfaction of a descriptive condition ϕ at W_A whichever W_P plays the role as W_A . We get *Two-Dimensionalist Reference*:

(TDR) N refers to o at W_P qua counterfactual iff o satisfies ϕ at W_P qua actual.

In particular, 'Julius' refers to an individual i at a counterfactual W_P iff i satisfies 'the inventor of the zip' at W_A whichever W_P turns out to be actual. Similarly, when it comes to reference at W_A , *Two-Dimensionalism* rules that:

(W_A -R*) N refers to o at W_P qua actual iff o satisfies ϕ at W_P qua actual.

When I said in Sec. 2.3 that Descriptivism and Referentialism alike endorsed (W_A -R), I took it for granted that ' W_A ' was a fixed world. But just as in (TDR), we must now allow for the possibility that we consider a different W_P as a candidate for W_A . So, if a term like 'Julius' contains a hidden actuality-operator, then its reference will obviously be sensitive as to which W_P we consider as actual. To consider another W_P as actual, is to consider W_P as a possible context of utterance or acquisition. So, given that 'actually' is an indexical expression, the reference of 'Julius' will be sensitive to variations in W_P qua actual. But even if 'Julius' did not contain 'actually', its reference would still be sensitive to variation in which W_P is considered as actual given that such hypotheses involve considering our stipulation (S^*) as made in those W_P . (S^*) fixes the reference of 'Julius' to whoever invented the zip in W_A , but had (S^*) been made in a W_P in which someone else invented the zip, then 'Julius' would have referred to him. So, giving up on 'actually' need not rob 'Julius' of its Two-Dimensional semantics.²³

²³ The point is thus that a term N is two-dimensional iff it is possible that N refers to an object at a W_P qua actual but to a different object at the same W_P qua counterfactual. In other words, N is two-dimensional iff the truth-value of a sentence containing N has a double world-dependency: it depends on which W_P we take as

4.4. Two Grades of Understanding

Suppose now that we model the semantic functioning of 'water' on 'Julius'. Our stipulation:

(S) Let 'water' refer to whatever stuff has watery properties $P_1, P_2, P_3 \dots P_n$,

both fixes the reference of 'water' to refer rigidly to the kind common to the watery exemplars of our acquaintance in W_P considered as actual, but also confers the descriptive content of 'the watery stuff of our acquaintance' on 'water'. Had (S) been carried out in a different W_P , 'water' would have been a rigid designator of the watery stuff of our acquaintance in that W_P , but given that (S) is performed in W_A , 'water' rigidly picks out the watery stuff of our acquaintance in W_A . 'Water' can thus be seen as shorthand for 'the watery stuff of our actual acquaintance'. That way, 'water' has associated a cluster of descriptive content and yet is neither susceptible to the Twin Earth Argument nor to the Modal Argument since we have built into the cluster conditions of rigidity and causality.

actual and which W_P we take as counterfactual. The fact that singular terms are two-dimensional should be common ground between Descriptivism and Referentialism. Both camps could endorse (TDR) and (W_A -R*). Note, for instance, that (TDR) and (W_A -R*) simply amounts to (TDR*) N refers to o at W_P iff o satisfies ϕ at W_P qua actual, which differ from (RR) only in that the latter uses a fixed notion of ' W_A '. There would, however, still be disagreement between them as to what makes it the case that N is two-dimensional. The Descriptivist insists that it is in virtue of descriptive content that N has this property, whereas the Referentialist maintains that N has this property if there is a reference-fixing description. E.g. Stalnaker's [1999a] *propositional concept* is two-dimensional since it is defined as a function from W_P qua actual into a proposition which in turn is a function from a W_P qua counterfactual into a truth-value. But Stalnaker's propositional concept is certainly not descriptive in the sense I use that notion. Unlike Kaplan's [1989] *character* which is a two-dimensional concept associated with an expression-type as its meaning, a propositional concept is a context-sensitive property of expression-tokens; cf. also Lewis [1980]. What is important to bear in mind is that the fact that N is two-dimensional, and thus can be represented by a matrix, is in itself no argument that N has descriptive content. The contention is only that *if* N has descriptive content, *then* we can use the two-dimensional framework to represent its semantics.

'Water' is thus a *Two-Dimensional* notion.²⁴ Let W_1 be Earth where H_2O is the watery stuff of our acquaintance and let W_2 be Twin Earth where XYZ is the watery stuff of our acquaintance. We can then depict utterances of 'water' by the matrix in fig. 5:

fig. 5

	W_1	W_2
W_1	H_2O	H_2O
W_2	XYZ	XYZ

The descriptive content of 'water' is what determines the reference of 'water' in a W_P , i.e. when we consider W_P as actual. Considering a W_P as actual involved in the case of indexicals considering W_P as a possible *context of utterance*: if I speak, then 'I' refers to me, but if you had spoken, then 'I' would have referred to you, etc. But in the case of natural kind terms, a shift in context of utterance is clearly not enough to bring about a change in reference. The so-called switching arguments (cf. Sec. 1.4) emphasised that only after *some time* on Twin Earth would an Earthling acquire twin-water concepts. When we consider the possibility that another W_P is actual, we must thus take W_P to be a possible *context of acquisition*. We can then represent the descriptive content of 'water' as a function from W_P qua context of acquisition into a referent *in* W_P , i.e. as a function that tells us, for each W_P , what 'water' refers to in W_P . But this is exactly what the *diagonal proposition* depicts: 'water' refers to H_2O in W_1 and 'water' refers to XYZ in W_2 . The *horizontal proposition*, however, is a function from a W_P considered counterfactual into a referent *of* that W_P . In W_1 'water' refers to H_2O of W_1 and of W_2 . Which kind of propositional content is represented by the horizontal proposition such that on Earth 'water' refers to H_2O of all W_P ? Well, consider a sentence like:

(25) Water is wet,

²⁴ Cf. also Davies & Humberstone [1981a, pp. 18-20], Chalmers [1996, p. 59], Jackson [1998a, pp. 212-4] and Byrne [2000a; 2000b].

and suppose that there is a minor amount of non-watery H_2O on Twin Earth (W_2) and a minor amount of non-watery XYZ on Earth (W_1). We can then represent utterances of (25) by fig. 6:

fig. 6	W_1	W_2
W_1	True	False
W_2	False	True

Now ask: what are the truth-conditions of (25)? It depends. In Sec. 2.6 I talked about W_A -*truth-conditions*, i.e. the conditions under which a sentence is true in W_A . Now extend this notion to cover the conditions under which a sentence is true at W_P considered as actual. Then we can ask: what are the W_A -truth-conditions of (25), i.e. the conditions under which (25) is true-in all W_P ? The Descriptivist will then insist that the causally constrained descriptive content associated with 'water' yields perfectly adequate W_A -truth-conditions: 'water is wet' is true-in W_P iff the watery stuff of our acquaintance in W_P is wet in W_P . Causally constrained descriptive content gives the right truth-conditions for all W_P considered as actual. (25) is true-in W_1 iff the watery stuff of our acquaintance in W_1 is wet in W_1 , and (25) is true-in W_2 iff the watery stuff of our acquaintance in W_2 is wet in W_2 . (25) is associated with the *same* descriptive proposition in all W_P and this proposition gives the condition under which (25) is true-in all W_P . So, the claim is that descriptive truth-conditions are what give W_A -truth-conditions. Formally:

Descriptive truth-conditions: 'N is F' is true-in W_P iff ϕ in W_P is F in W_P ,

where N is a singular term associated with descriptive content ϕ . Descriptive truth-conditions reflect the fact that even if N is modally rigid and ϕ modally flexible, 'N is F' and ' ϕ is F' will have identical truth-conditions not only at W_A , but at every W_P when considered as actual; provided the descriptive content of N is given by ϕ . They say the same thing about how W_A is no matter which W_P turns out to be actual, that is, they are

made true by the same state of affairs in W_P qua actual. It is simply unintelligible that evidence in W_A should bear differently on the truth-value of the two sentences. The only difference between them arises when we turn to counterfactual W_P , i.e. when they are prefixed with modal idioms.

However, as saw in Sec. 2.3, the Referentialist insists that utterances of sentences have nothing but singular truth-conditions. Whether we consider W_A or counterfactual W_P , she holds, in general:

Singular truth-conditions: 'N is F' is true at W_P iff o is F in W_P ,

where o is the referent of N, fixed by unique satisfaction of some descriptive condition ϕ . As we saw, singular truth-conditions get things right with respect to W_A and counterfactual W_P . But they fall short of accommodating W_A -truth-conditions in the expanded sense of that notion. Singular truth-conditions do not say under what conditions an utterance of 'N is F' is true had another W_P been actual. What the Referentialist must do, it seems to me, is to come up with an explanation in terms of reference-fixing. Take (25) 'water is wet'. In W_1 'water is wet' is true iff H_2O is wet, but had W_2 been actual, the reference of 'water' would have been fixed by (S) to pick out XYZ, and so 'water is wet' would have been true iff XYZ is wet. So, even if the Referentialist refuses to accept that stipulation (S) confers descriptive content on 'water', she must assign a significant role to (S) when accounting for W_A -truth-conditions of sentences containing 'water'.

What about the conditions under which a sentence like (25) is true-of W_P , i.e. what I in Sec. 2.6 called *W_P -truth-conditions*? I argued that there were two equally adequate ways of giving the truth-conditions of a sentence with respect to counterfactual W_P . Given that the Referentialist endorses (Singular-Rigidity), she would insist that such sentences have singular truth-conditions: on Earth, 'water is wet' is true-of W_P iff H_2O is wet in W_P , and on Twin Earth, 'water is wet' is true-of W_P iff XYZ is wet in W_P . Once we have fixed a W_P qua actual, the singular content of 'water' gives the conditions under which (25) is true-

of every W_P , i.e. at every counterfactual W_P . Singular truth-conditions reflect the fact that 'N is F' and ' ϕ is F' have different truth-conditions when it comes to counterfactual W_P , and so that ϕ cannot exhaust the propositional content of N. The response, on behalf of Descriptivism, was to rigidify ϕ such that we would get the correct truth-conditions even at counterfactual W_P : on Earth, 'water is wet' is true-of W_P iff the watery stuff of our acquaintance *on Earth* is wet in W_P . Formally:

Rigidified Descriptive truth-conditions: 'N is F' is true-of W_P iff @ ϕ is F in W_P .

Both singular and rigidified descriptive truth-conditions respect the fact that (25) is true-in W_1 , but is, in W_1 , false-of W_2 , and that (25) is true-in W_2 , but is, in W_2 , false-of W_1 .

Now, in Sec. 2.6 I distinguished between two grades of understanding. I said that if a competent speaker grasps the descriptive truth-conditions of a sentence, then she knows the conditions under which the sentence is true in W_A . But it is now clear that if she knows the descriptive content of a sentence, then she knows not only the conditions under which the sentence is true given the way W_A has turned out, she also knows the conditions under which the sentence is true had another W_P turned out to be actual. We shall now say that a speaker who grasps the *causally constrained descriptive truth-conditions of a sentence has a grade-one understanding* of it. To have a grade-one understanding is to have knowledge of W_A -truth-conditions: it is to know under what conditions an utterance of a sentence is in fact true and under what conditions the utterance would have been true had a different W_P been W_A .²⁵ To have a *grade-one understanding is thus to know the diagonal proposition*. In the case of (25) 'water is wet', it is to know that (25) is true-in W_1 iff the watery stuff of our acquaintance in W_1 is wet in W_1 ; (25) is true-in W_2 iff the watery stuff of our acquaintance in W_2 is wet in W_2 . So, having a grade-one understanding enables one to determine how a set of singular truth-conditions depend on which W_P is considered as actual. If H_2O is the watery stuff of our acquaintance in W_1 , then (25) is true-in W_1 iff H_2O

²⁵ The same idea is explicit in Dummett [1981, p. 568].

is wet in W_1 ; if XYZ is the watery stuff of our acquaintance in W_2 , then (25) is true-in W_2 iff XYZ is wet in W_2 .

Although a speaker who has a grade-one understanding of a sentence has a fairly extensive grasp of its assertoric use, her competence is essentially incomplete. She does not know how the sentence embeds inside the scope of a modal operator as she is ignorant of the modal properties of its component expressions. In order to fully understand a sentence, knowledge of its rigidified descriptive truth-conditions is mandated, e.g. whether it contains rigid designators. We shall say that a speaker who grasps the *rigidified, causally constrained, descriptive truth-conditions of a sentence* has a *grade-two understanding* of it. To have a grade-two understanding of (25) is thus to know that (25) is true-of W_1 iff the actual watery stuff of our acquaintance is wet in W_1 ; (25) is true-of W_2 iff the actual watery stuff of our acquaintance is wet in W_2 . Here 'actual' refers to W_A or a W_P if considered as actual. So, having a grade-two understanding enables one to determine how a set of singular truth-conditions depend on which W_P is considered as actual. If H_2O is the actual watery stuff of our acquaintance, then (25) is true-of W_P iff H_2O is wet in W_P ; if XYZ is the actual watery stuff of our acquaintance, then (25) is true-of W_P iff XYZ is wet in W_P . This highlights the fact that to have a *grade-two understanding* is not to know a *horizontal proposition*. To know a horizontal proposition requires that one affirms an antecedent in one of these conditionals, but one is not able to do that just by knowing the rigidified descriptive content of 'water'; one must also know the reference of 'water'. To know a horizontal proposition is to know a function from counterfactual W_P into truth-values, but given that such truth-values also depend on which W_P is considered as actual, one must first know the relevant empirical facts about that W_P . What does amount to knowledge of a horizontal proposition is knowledge of singular content. Only someone who *knows the singular truth-conditions* knows a *horizontal proposition*.²⁶

²⁶ In Jackson [1998a] and Chalmers [1996], the W_A -truth-conditions are identified with the diagonal proposition - what they call the A(ctual)-intension and the primary intension respectively, and the W_P -truth-

Suppose S is an Earthling who has a grade-two understanding of (25), and is asked whether (25) is true-of Twin Earth. All S can say is that (25) is true-of Twin Earth iff the watery stuff of her acquaintance on Earth is wet on Twin Earth. What S knows is the conditional that if XYZ is the watery stuff of her acquaintance on Earth, then (25) is true-of Twin Earth; otherwise not. So, S does not know the horizontal proposition since she does not know the truth-values of (25) at various counterfactual W_P . But the Referentialist will now insist that to know how the singular proposition expressed by a sentence is dependent on which W_P is considered as actual is not to fully grasp what an utterance of the sentence expresses. To know that one must know the singular proposition expressed. Imagine another Earthling R who knows that H_2O is the watery stuff of our acquaintance on Earth. R is able to detach the antecedent in our conditional, and so knows that (25) is true-of W_P iff H_2O is wet in W_P . R thus has a much better grasp of the truth-conditions of (25) at counterfactual W_P . Unlike S, R knows the truth-values of (25) at various counterfactual W_P . So, the claim is that only knowledge of singular truth-conditions constitutes a full grasp of the truth-conditions of (25).

It is plain that R knows more about the world, e.g. about micro-physics, than S does, but the issue is whether R is best described as having a better linguistic understanding than S has. The Descriptivist will resist the temptation to describe R as having knowledge of propositional content that S lacks; what R has is knowledge of reference or truth. R knows that 'water' rigidly picks out H_2O on Earth, and so knows that an Earthly utterance of (25) is false-of Twin Earth where XYZ is wet. But, the Descriptivist's thought goes, one need not know the reference of 'water' in order to understand utterances of sentences

conditions are identified with the horizontal proposition - what they call the C(ounterfactual)-intension and the secondary intension respectively. Since I hold that propositions are *structured* entities, and hence more fine-grained than sets of worlds or functions from worlds into truth-values, descriptive and singular contents should not be *identified* with the diagonal and horizontal propositions respectively: the former *determine* the latter but not vice versa. The risk is that if, say, descriptive content is deflated in this way, then we waive any attempt to give a substantial account of which properties constitute such content.

containing 'water'. That is, one need not know that H_2O is the watery stuff of our acquaintance in order to understand such utterances - a sufficiently comprehensive grasp of the associated descriptive properties will suffice. In the case of *atomic* sentences, all understanding takes is knowledge that 'water' picks out the watery stuff of our acquaintance, whatever that is, and in the case of *modalized* sentences, all understanding takes is knowledge that 'water' picks out the *actual* watery stuff of our acquaintance, whatever that is. To understand an atomic sentence is to know its W_A -truth-conditions, and to understand a sentence containing modal idioms is to know its W_P -truth-conditions. In neither case is knowledge of singular content mandated.

Here is a putative counterexample. Suppose Q says 'I have a headache'. S, who has a grade-one understanding, has a very poor grasp of what Q said. S neither knows who the speaker is nor that 'I' is a rigid designator. So all S knows is that if Q uttered 'I have a headache', then 'I have a headache' is true-in W_P iff Q has a headache in W_P . In words, she knows that someone said that (s)he had a headache. Now consider R who has a grade-two understanding. R knows that 'I' is a rigid designator, but is still unaware who the speaker is. So, R knows that if Q uttered 'I have a headache', then 'I have a headache' is true-of W_P iff Q has a headache in W_P . The claim is that S and R are in the same predicament: they do not understand what Q said because they do not know who spoke. In such cases where the non-singular content is next to uninformative - a linguistic rule - and where the singular content is easily read off from the context of utterance - the speaker - we tend to demand that understanding proper be of singular content.

It is true that we commonly assign singular truth-conditions to utterances of sentences containing indexicals and demonstratives. But even in such cases do sentences have a kind of *default* descriptive truth-conditions. If one is unaware who the speaker is, then one still has a partial understanding of what was said: 'I have a headache' is true iff the speaker of the context has a headache. Moreover, what a speaker knows when she understands the sentence-type 'I have a headache' is such descriptive truth-conditions; it is

just that in the sense of fully understanding what is said by a token of that sentence, she must also know who 'T' picks out in the context of utterance. In any case, the Descriptivist can insist that sentences containing indexicals and demonstratives are exceptional. There is no parallel to sentences containing natural kind terms where the descriptive content is very informative and where the singular content takes substantial empirical investigation. In such cases, knowledge of - rigidified - descriptive content is both necessary and sufficient.

The difference between a speaker R who knows rigidified descriptive truth-conditions, and a speaker T who knows singular truth-conditions, is thus that T has empirical knowledge about the context of utterance or acquisition that R lacks. T knows which W_P is actual such that, in the case of 'water', he knows *which kind of stuff* 'water' refers to in W_A . T has De Re knowledge of reference that S lacks, that is, T knows of H_2O that 'water' refers to it. Such empirical knowledge enables T to determine reference or truth at various counterfactual W_P . Given that T knows that the watery stuff on Twin Earth is XYZ, T knows that Earthly tokens of 'water' do not refer to XYZ on Twin Earth. T thus knows that Earthly tokens of (25) 'water is wet' are false-of Twin Earth. R, on the other hand, merely knows that such tokens are true-of Twin Earth iff the actual watery stuff of our acquaintance is wet on Twin Earth. In illustration:

(A) Knowledge of rigidified descriptive content + knowledge of W_P qua actual \rightarrow knowledge of singular content + knowledge of W_P qua counterfactual \rightarrow Knowledge of reference/truth-value.

What R has is thus knowledge of *object-independent* truth-conditions: (25) is true-of W_P iff the watery stuff of our actual acquaintance is wet in W_P . T, however, has knowledge of *object-dependent* truth-conditions: (25) is true-of W_P iff H_2O is wet in W_P . Or if we take Twin Earth as our context of acquisition, then (25) is true-of W_P iff XYZ is wet in W_P . This means that (25) expresses *different* singular propositions, but the *same* descriptive proposition, in different W_P qua actual. So, the Two-Dimensional analysis emphasises that

according to Descriptivism (cf. Sec. 2.2), 'water' is *constant in content* across W_P qua actual: 'water' is associated with the propositional content <the watery stuff> both on Earth and on Twin Earth. According to Referentialism (cf. Sec. 2.3), however, 'water' is *variant in content* across W_P qua actual: on Earth tokens of 'water' express < H_2O >, and on Twin Earth tokens of 'water' express <XYZ>.²⁷ Both Descriptivism and Referentialism agree about reference. 'Water' is *world-relative in reference* in the sense that the reference of 'water' is sensitive to variation in contexts of acquisition: 'water' refers to H_2O on Earth qua actual, but 'water' refers to XYZ on Twin Earth qua actual. There is also a sense in which 'water' is *world-bound in reference*: given that 'water' refers to H_2O on Earth, 'water' refers to H_2O of all W_P , and given that 'water' refers to XYZ on Twin Earth, 'water' refers to XYZ of all W_P . That is, both views can agree that 'water' is a rigid designator.

Let me finish this Section by returning to the *Semantical Argument* against Descriptivism, as promised in fn. 35, Sec. 2.5.²⁸ Although the Semantical Argument is akin to the Modal Argument, it highlights an important difference. The Modal Argument, remember, aimed at showing that if the content of, say, 'Aristotle' is given by 'the teacher of Alexander', then 'Aristotle' refers to Plato at a counterfactual W_P in which Plato taught Alexander, but 'Aristotle' refers to Aristotle at all counterfactual W_P . The Semantical Argument, however, purports to show that if the content of 'Aristotle' is given by 'the teacher of Alexander', then *if Plato had taught Alexander, then 'Aristotle' would refer to Plato*, but 'Aristotle' would refer to Aristotle were such counterfactual circumstances to obtain. So, the Modal Argument is concerned with what a term refers to at counterfactual

²⁷ This is explicit in Burge [1982, p. 105]: "The fact that the Twin-Earthians apply 'water' to XYZ is not a reflection of a shift in extension of an indexical expression with a fixed linguistic (English) meaning, but of a shift in meaning between one language, and linguistic community, and another. Any expression, indexical or not, can undergo such 'shifts', as a mere consequence of the conventionality of language."

²⁸ Versions of the Semantical Argument can be found in Kripke [1980, pp. 83-92], Donnellan [1972, pp. 373-75] and Salmon [1982, pp. 29-31].

W_P. The Semantical Argument is concerned with what a term would refer to had certain circumstances obtained.

In light of Two-Dimensionalism, it is not difficult to respond to this Argument. First of all, no contemporary Descriptivist holds that a singular term is simply an abbreviation of a single modally flexible definite description. According to Causal Rigidified Descriptivism, singular terms have associated with them clusters of descriptive as well as non-descriptive properties. Suppose one of the properties associated with 'Aristotle' is the property of being the individual who actually taught Alexander. What then if it is later discovered that in actual fact Plato taught Alexander? Then I could call upon other properties: the author of *De Anima*. But what if Plato also actually wrote *De Anima*? Then I could appeal to other-dependent identifying information: the man called 'Aristotle' by my speech community. But what if they were all wrong? Then I must resort to the causal chain originating in my use of 'Aristotle': the man causally responsible for my use of 'Aristotle'. But what if Plato satisfies even that description? Well, then 'Aristotle' refers to Plato! If everything true of Aristotle were true of Plato, then 'Aristotle' would simply refer to Plato. Were such circumstances to obtain, unlikely as they are, then we would have to admit that we thought we referred to a certain individual by 'Aristotle' when all along we had been referring to someone else. What we are considering is thus what it would be true to say had another W_P been actual, i.e. were we to use 'Aristotle' under circumstances in which everything true of Aristotle is true of Plato, we would be referring to Plato.²⁹

Compare with 'water'. Imagine circumstances in which the watery stuff that we actually and causally interact with turns out to have the micro-structure XYZ. The question is: were such circumstances to obtain, would 'water' refer to XYZ? The answer is yes. What we are considering is Twin Earth as a candidate for W_A, but if Twin Earth is considered as actual, i.e. as a possible context of acquisition, then our tokens of 'water' refer to XYZ.

²⁹ What must be shown to sustain the Semantical Argument is that singular terms are what Evans called 'deeply rigid', i.e. that they refer to the same object not only of all W_P, but also in all W_P; cf. fn. 20.

Although highly improbable, it is not unimaginable that physical science one day discovers that we have made a huge mistake in believing that the watery stuff that surrounds us has the micro-structure H_2O . To be clear, what is conceivable are circumstances under which an utterance of 'water is XYZ' would have been true. That, however, is not to conceive of circumstances under which water is XYZ: if water is H_2O , then it is necessary that water is H_2O .³⁰

4.5. The Contingent A priori

The foregoing Two-Dimensional analysis suggests that the two propositions that can be expressed by an utterance of a simple sentence have different epistemic status. To know the singular proposition expressed is to know an object-dependent truth-condition, e.g. it is to know that, on Earth, (25) 'water is wet' is true-of Twin Earth iff H_2O is wet on Twin Earth. To know the singular proposition expressed is thus to have relevant knowledge about the context of utterance or acquisition that no amount of semantic introspection could result in. In the case of 'water', it is to have relevant knowledge of micro-physics that one could only acquire by adequate empirical means. To know the singular proposition expressed is thus to have *a posteriori* knowledge about the W_P that is in fact actual, i.e. W_A . By contrast, one can know the descriptive proposition associated with (25) without knowing the relevant micro-physical facts. Descriptive propositions are object-independent since they do not depend on the nature of the W_A . To know the descriptive proposition of (25) is to know the truth-conditions of (25) at every W_P qua actual, and so does not depend on which W_P is in fact actual. This is what the diagonal proposition reflects. To know the descriptive

³⁰ This means that there are really *three ways of taking Twin Earth*: (i) Twin Earth as a remote planet in W_A ; (ii) Twin Earth as a counterfactual W_P ; (iii) Twin Earth as a way W_A might have turned out; cf. also Jackson [1998b, pp. 212-14; 2000, pp. 330-1]. See also Sec. 4.6.

proposition is thus to have knowledge of truth-conditions which is independent of empirical knowledge of W_A , and so is to have *a priori* knowledge.

As we saw in Sec. 4.4, the claim that there is an *a priori* and an *a posteriori* aspect to understanding is most plausible when it comes to indexicals. Suppose S says in the presence of R: 'I am here'. What they both know *a priori* are certain linguistic rules that take 'I' and 'here' to the utterer and the place of the context. S wishes to convey the proposition that he is at place *p*, and he uses rules governing 'I' and 'here' which he assumes R is familiar with. If R is to grasp the singular proposition expressed, then R need not only be aware of these rules, R must also acquire the relevant perceptual knowledge that S is the contextually salient speaker located at *p*.

The reason we think that R must know the singular proposition expressed in order to fully understand what S said, is that the descriptive content associated with indexicals is very poor and the singular content is easy to get at. But when it comes to non-indexical singular terms where the descriptive content is rich and the singular content harder to get at, there is no reason to think that full understanding must require knowledge of singular content. In the latter cases, there would therefore not seem to be the same epistemic bifurcation of understanding; or so Descriptivism has it. This is what the *Epistemological Argument* takes issue with. It goes like this: take 'Aristotle'. If 'Aristotle' were simply a shorthand for 'the teacher of Alexander', then the sentence:

(26) If Aristotle exists, then Aristotle is the teacher of Alexander,

would express an *a priori* knowable proposition since semantically equivalent to:

(27) If the teacher of Alexander exists, then the teacher of Alexander is the teacher of Alexander,

which expresses an *a priori* knowable proposition. But (26) does not express an *a priori* knowable proposition since it was an empirical discovery who taught Alexander. So, 'Aristotle' cannot be short for 'the teacher of Alexander'.³¹

³¹ Cf. Soames [1998, p. 1]. See also Salmon [1982, pp. 27-9] and Block & Stalnaker [1999].

As stated, the Argument looks sound. It is, to be sure, a version of the CP-Argument as explained in Sec. 2.4 which assumes that we can use Leibniz Law to show that a difference in epistemic properties entails a difference in propositional content. But such an assumption would not seem to beg the question against Descriptivism according to which propositional content is individuated by its cognitive properties. So, is there any other way to fault the Argument? Well, one lesson to be learned so far is that no Referentialist argument will carry conviction if it relies on an implausible version of Descriptivism. Thus no present-day Descriptivist holds this single-description view when it comes to ordinary proper names and natural kind terms. Only the artificiality of 'Julius' makes an exception. This name has no use prior to the stipulation (S*) which essentially introduces a modally rigid referring name for the inventor of the zip. Henceforth, full competence with 'Julius' must include knowledge of (S*). So, it would seem to be a priori knowable that if Julius exists, then Julius is the inventor of the zip. But the same is presumably not true of 'Aristotle'. According to Descriptivism, 'Aristotle' has associated with it a cluster of properties: the actual teacher of Alexander, the actual author of *De Anima*, the causal source of our use of 'Aristotle', the individual referred to by the experts' use of 'Aristotle', etc. We have seen cases where it is indispensable that some of these properties involve causality and rigidity. But we can also imagine cases where 'Aristotle' succeeds in referring despite the fact that no single individual satisfies all and only the associated properties. If it turns out that Plato taught Alexander, then 'Aristotle' will still refer to Aristotle in W_A provided he has all the other properties. We should not think of the cluster properties as necessary and sufficient conditions for being picked out by 'Aristotle'. What Descriptivism claims is merely that it is (i) *a priori* that 'Aristotle' refers to whoever has (ii) *sufficiently many* of the associated properties (iii) *in* W_A . We know that when we consider counterfactual and Twin Earth cases, modal and causal properties carry great weight. Given that 'Aristotle' is a rigid designator, we can allow for counterfactual W_P in which Aristotle has - next to - none of the properties he has in W_A . What is a priori is not that the referent

of 'Aristotle' has the descriptive properties in counterfactual W_P , but that he has enough of them in W_A . How many that is and which have the greater weight in the vote is somewhat indeterminate and probably settled by the linguistic community from case to case. What if it turns out that half of what we thought we knew about Aristotle was actually true of Plato? Well, if that is how W_A is, then presumably 'Aristotle' is empty. Or, maybe after having balanced the properties, 'Aristotle' still refers to Aristotle. In any case, the point is that although, for each descriptive property, it is possible that Aristotle lacks that property in W_A and yet is still the referent of 'Aristotle', it is not possible that Aristotle lacks all descriptive properties in W_A and yet is the referent of 'Aristotle'.³² And the reason we know that a priori is, as the Semantical Argument emphasised, that we can engage in a priori reflection on what 'Aristotle' would refer to had various circumstances obtained: had Plato been the teacher of Alexander, then 'Aristotle' would still refer to Aristotle, but had *everything* actually true of Aristotle been true of Plato, then 'Aristotle' would refer to Plato. Knowledge of what 'Aristotle' would refer to had the way W_A is been different in various ways is a priori since it is independent of the way W_A in fact is.

Similarly in the case of 'water'. The stipulation (S) tells us that 'water' is set up as a rigid designator of whatever stuff has *sufficiently many* of the watery properties in W_A . We can undoubtedly imagine something being water in W_A without having all the purely descriptive properties, and something having some of those characteristic properties in W_A without being water. As Putnam [1970] reminded us of, there are non-striped, non-yellow tigers and striped, yellow non-tigers even in W_A . There may be vagueness or elasticity in how many of the watery properties something must have in order to be water, and it is also possible that the list of watery properties is open-ended. What seems impossible is merely that something is water in W_A and yet lacks *every* characteristic of water, or that something

³² Formally, there is no move from $\forall P(Pa \rightarrow \Diamond \sim Pa)$ to $\Diamond \forall P(Pa \rightarrow \sim Pa)$.

has *all* the watery indicators in W_A and yet is not water.³³ In this respect 'water' is not like 'Julius': it is chronologically incorrect to say that (S) introduces 'water' into the language with a fixed descriptive content. What (S) reflects is the kind of *role* 'water' plays in our community and to spell out that role is to register a set of core, stereotypical beliefs about water which prevail in the community; that water is the predominant potable liquid that fills the oceans, falls from the sky, etc.; in short the watery stuff.³⁴ The sense in which 'water' is akin to 'Julius' is that there will be community-wide consensus on the range of properties that make up its a priori knowable descriptive content. Consider the sentence:

(28) If water exists, then water is watery.

The descriptive content of 'water' is encapsulated by 'the watery stuff' which records our stereotypical beliefs about water. This means that (28) is a priori in the sense that understanding (28) suffices for knowing that it is true. Someone who fully grasps the propositional contents in (28) can come to know its truth-value just by reflection, which of course is not to say that she can come to grasp those contents just by reflection; concept-acquisition standardly requires experiential information gathered from observation or

³³ Let me stress that I have no firm intuitions about how to describe various possible cases. In Sec. 1.2 I said that water could not have turned out to be like jade. What I meant was that if the extension of 'water' on Earth is H_2O , then my use of 'water' would not pick out XYZ on a remote planet in W_A . But what if Earth turns out to have a 50/50 mix of H_2O and XYZ? Then probably the extension of 'water' would comprise both. So, what if Earth turns out to have a 95/5 mix of H_2O and XYZ? Then probably 'water' would refer only to H_2O ; cf. also Chalmers [1996, p. 58]. We can think of this *cluster* of watery properties as "...a disjunction of all conjunctions of *most* of them", as Lewis [1991, p. 208] has put it in a different context; cf. also Jackson [1998a, p. 35].

³⁴ Lewis [1999b] and Jackson [1998a] have argued that *conceptual analysis* shows that theoretical terms like 'water' are *cluster or functional role concepts*. That is, conceptual analysis shows that water is the kind of stuff that typically occupies the water-role. Spelling out the water-role is a matter of saying what a competent speaker knows when she understands 'water'-sentences. This type of conceptual analysis should be distinguished from the traditional project of pinning down sufficient and necessary conditions for something to be water. In spelling out 'the water-role' one must draw on - implicit and explicit - common-sense knowledge about the stereotypical characteristics associated with water.

passed on via testimony. We know that no matter how W_A turns out, nothing can be water in it unless it has *sufficiently* many of the watery properties, and we know this a priori since it is arrived at just by reflection on what we would say had various W_P been actual and is thus knowledge which is not dependent on the way W_A in fact is. (28) can be pictured by fig. 6 and its diagonal proposition contains nothing but truths. It thus records the fact that (28) is true at every W_P considered as actual and so can be known to be true at W_A without knowing which W_P that is. *What is a priori is what is true no matter how W_A turns out.*³⁵

Now, when the Epistemological Argument is advanced, it is frequently supposed that the reason a sentence like (28) cannot express an a priori knowable proposition is that it is not necessary that water is watery.³⁶ The line of reasoning is this: if it is *a priori* that if water exists, then water is the watery stuff, then it must be *analytic* that if water exists, then water is the watery stuff. But analytic truths are necessary, so it must be *necessary* that if water exists, then water is the watery stuff. Yet it is clearly possible that water, i.e. H_2O , exists, but is not the watery stuff. After all, 'water' and 'the watery stuff' have distinct modal properties. So, it cannot be a priori that if water exists, then water is the watery stuff.

I think the objection confuses a priority with necessity. (28) is an example of an a priori contingency, but it is obviously worth dwelling on how this is supposed to work. I suspect there are two explanations of how such statements are possible. Here is the first. Take an utterance by S of 'I am here now'. S is at place p at time t, and so expresses the singular proposition that S is at p at t. This singular proposition is contingent since S might have been somewhere else than where he is when that token occurs, and only a posteriori knowable since it requires empirical knowledge of relevant features of the context. The sentence 'I am here now' is, however, also associated with the descriptive proposition that the speaker of the context is at the place and time of the context. And that proposition is

³⁵ When I said that fig. 4 represented (24) and fig. 6 represented (25) I assumed that 'Julius' and 'water' had a referent in each W_P . We need the proviso on existence in (28) since it is not a priori knowable that 'water' is not empty.

³⁶ This is explicit in Salmon [1982] and Block & Stalnaker [1999].

both necessary and a priori knowable. There is no possibility of reference-failure and no need to check the speaker, time, and place of the context to assure oneself that the speaker is at the time and place of the context; whoever, whenever, or wherever that is. It follows from the linguistic meaning of the indexicals that the speaker is whoever utters a certain sentence when located at the place and time of the context. What this means is that the diagonal proposition for 'I am here now' consists solely of truths: its descriptive content is such that it cannot be used to express a false singular proposition.³⁷ But its horizontal proposition consists of both truths and falsehoods: it expresses a contingent singular proposition in each context of utterance. 'I am here now' is thus true-in every W_P , but is, at each W_P qua actual, false-of some W_P . So, in the case of 'I am here now' no single proposition is both contingent and a priori. Yet we may think of the sentence-token as an example of an a priori contingency. The same analysis applies to (28). That sentence is associated with the descriptive proposition:

(29) If the watery stuff exists, then the watery stuff is watery,

which is both necessary and a priori knowable. (29) is what someone knows when she understands (28). She knows the diagonal proposition, i.e. that (28) is true no matter how W_A turns out. (28) cannot be used to make a false assertion. But it is a familiar point that *to know (a priori) that a sentence expresses a truth is not to know (a priori) the truth of what it expresses*.³⁸ And what someone knows when she understands (28) is merely that its descriptive content is such that (28) expresses a truth in every W_P qua actual, that is, its diagonal proposition contains nothing but truths. In order to know the truth of what (28) expresses on Earth, she would have to know the singular proposition:

³⁷ There are tricky counterexamples. Suppose I am unaware that I am lost in the Scottish Highlands. While pointing at my map, I express a falsehood by 'I am here now'. Or suppose I call from work to leave a - true - message on my answering-machine at home: 'I am not here right now...'. In such cases, the context of utterance - where I am on the hill or where I am at work - and the circumstance of evaluation - where I point to on the map or where I am at home - come apart in the very act of uttering the sentences.

³⁸ Cf. Donnellan [1979], Schiffer [1978] and Salmon [1986].

(30) If H_2O exists, then H_2O is watery,

but she can only know that proposition once the appropriate micro-physical knowledge is in. However, (30) is contingent: there is a counterfactual W_P in which H_2O does not have the watery properties. This records the fact that (28) has horizontal propositions which consist of both truths and falsehoods: (28) is true-in every W_P , but is, at each W_P qua actual, false-of some W_P . So, there is no single proposition associated with, or expressed by, (28) which is both contingent and a priori.

Now, the Descriptivist may feel slightly uneasy about invoking singular content to account for the contingent a priori. So, here is a second explanation. What someone knows when she understands (28) is not only that it expresses a truth at every W_P when considered as actual, she also knows the truth of what it expresses. But the truth of what (28) expresses is not a singular proposition; it is the *rigidified* descriptive proposition:

(29*) If the watery stuff exists in W_A , then the actual watery stuff is watery.

Note first that the only difference between the two propositions in (29) and (29*) is that (29*) is rigidified. This means that whereas the proposition in (29) is trivially true at every counterfactual W_P , the proposition in (29*) is false at some counterfactual W_P . There are W_P in which the watery stuff in W_A , i.e. H_2O , is not watery. So, unlike (29), (29*) is a contingency. Nevertheless, (29*) is a priori knowable for much the same reason that (29) is. Being such that if the watery stuff exists, then the watery stuff is watery, is a property that *every* W_P has trivially. One can thus know a priori that the proposition in (29) is true at every W_P no matter how it is considered. Not so for (29*). One can only know whether the proposition in (29*) is true at a counterfactual W_P if one knows the relevant empirical facts about W_A . One cannot know whether the actual watery stuff is watery in a given counterfactual W_P unless one knows whether the watery stuff in W_P is identical to the watery stuff in W_A . But the point is that one can know a priori that the proposition in (29*) is true *at* W_A . Being such that if the watery stuff exists in W_A , then the actual watery stuff is watery, is a property that W_A has trivially; the rigidified proposition in (29*) will thus be

true at W_A no matter how it turns out. So, one can know just by conceptual reflection that the proposition in (29*) is true at W_A .

So, on this account, (28) can be assigned a single proposition which is both contingent and a priori. To know this proposition a priori, is however not to know any interesting empirical fact about W_A . The reason for this is, as Evans pointed out, that it is a harmless or a *superficial* sense of contingency. (29*) is not contingent in the *deep* sense of there being:

"... some state of affairs of which we can say both that had it not existed the statement would not have been true, and that it might not have existed."³⁹

Contrast (29*) with (30). (30) is a deep contingency: there is some state of affairs - the fact that H_2O is watery - which is such that had it not existed, (30) would not have been true, and that it might not have existed. Had H_2O not been watery, (30) would not have been true; e.g. (30) would not have been true had XYZ been watery. Moreover, it is true that H_2O might not have been watery, that is, there are counterfactual W_p in which it is not. But (29*) is not like (30). (29*) is not about any particular state of affairs. All it says is that the watery stuff in W_A , whatever it is, is watery. That of course is contingent in the sense that

³⁹ Cf. [1996, p. 212]. Evans' notion of superficial necessity corresponds to the ordinary \Box -sense of necessity, and his notion of deep necessity corresponds to the $\mathcal{J}\mathcal{F}@$ -sense of necessity from fn. 13. So, we can say that (28) is superficially contingent since ' \Box (28)' is false, i.e. (28) is false-of some p . But (28) is deeply necessary since ' $\mathcal{J}\mathcal{F}@$ (28)' is true, i.e. (28) is true-in every W_p . We can thus call $\mathcal{J}\mathcal{F}@$ an a priori truth operator: a sentence is a priori iff it expresses a truth at every W_p qua actual. Note two things: (i) It will in general be true that anything of the form $@s \leftrightarrow s$ is an a priori contingency. Suppose not, i.e. suppose that $\mathcal{Q}(@s \leftrightarrow s)$ is true. Then $\mathcal{Q}s \rightarrow \Box s$ is true by the K-axiom $\Box(p \rightarrow q) \rightarrow (\Box p \rightarrow \Box q)$, and so, $@s \leftrightarrow \Box s$ is true by the T-axiom $\Box p \rightarrow p$. But $@s$ may be true, while $\Box s$ false. It follows that $@s \leftrightarrow s$ is not a theorem in $S5@$. But $(@s \leftrightarrow s)$ is a priori since $\mathcal{J}\mathcal{F}@(s \leftrightarrow s)$ is true: $(@s \leftrightarrow s)$ is true-in every W_p . (ii) It is an interesting issue whether all deeply necessary truths are a priori. If one thinks, what I do not, that ordinary proper names are deeply rigid, i.e. that their reference is insensitive to variation in which W_p plays the role as actual, (cf. fn. 20, 29), then a sentence like "Superman = Clark Kent" is both superficially necessary and deeply necessary, but clearly not a priori. For discussion see Soames [1989, pp. 148-50], Davies & Humberstone [1981a, pp. 7-13].

there are counterfactual W_P in which the watery stuff in W_A is not watery. But (29*) is not contingent in the sense that its truth in W_A depends on some particular, contingent feature of W_A . It is the fact that H_2O is watery in W_A which makes (29*) true at W_A , but some other state of affairs could have made it true at W_A ; had XYZ been the watery stuff in W_A , then (29*) would have been true at W_A . In short, the sentence 'if H_2O exists, then H_2O is watery' is not true-in every W_P , but 'if the watery stuff exists in W_A , then the actual watery stuff is watery' is. So, to know (29*) a priori is not to know anything informative about empirical reality.⁴⁰

4.6. The Necessary A Posteriori

We have seen in the last Section how the Two-Dimensional framework offers a neat explanation of the contingent a priori; we shall now see how the necessary a posteriori is equally well handled. But before I get to that, it is worth going back to the Intentional Argument against Referentialism. That Argument, remember, aimed at showing that the propositional content of a singular term could not be exhausted by its referent due to failure of substitutivity in belief contexts. The question is whether we can run a parallel argument in knowledge contexts by substituting co-referring terms occurring inside the scope of an epistemic operator. I think so. We saw that Referentialism is committed to the Substitutivity Principle (PS) in belief contexts, so let us now define *Strict Referentialism* as the view that (PS) holds across the board: the sole semantic function of a singular term is to refer to an object such that no descriptive condition can enter into the propositional content of a sentence containing that term even if that sentence contains epistemic operators. Now, if we bear in mind the Superman legend, it would seem that:

(31) Lois Lane knows that Superman is Superman,
is true, while:

⁴⁰ For more discussion of the contingent a priori see Salmon [1988] and Kallestrup [2000].

(32) Lois Lane knows that Superman is Clark Kent,
is false. We can now argue as follows. If (31) and (32) differ in truth-value, then, by Leibniz Law, they should also express different propositions, since we agreed that propositional content is individuated by its truth-conditions. Yet the only difference between (31) and (32) is that the former contains 'Superman' where the latter contains 'Clark Kent'. So, given (Compositionality of Content), the propositional contents of 'Superman' and 'Clark Kent' must be different. But this contradicts Strict Referentialism.

This is no doubt too quick. The Strict Referentialist can invoke conversational implicatures to explain away the prevailing, but nevertheless misguided, intuition that (31) and (32) differ in truth-value. To simply build the falsity of (32) into the story is to beg the question against the Strict Referentialist. Now, maybe knowledge ascriptions resemble belief ascriptions in that both occasionally generate false implicatures. Anyway, it is obvious how to circumvent this move and strengthen the argument. It seems indisputable that:

(33) Lois Lane knows a priori that Superman is Superman,
is true, while:

(34) Lois Lane knows a priori that Superman is Clark Kent,
is false. That is, if there is a sense in which Lois Lane knows the identity, then that knowledge is clearly not reflectively accessible to her; after all she is disposed to dissent from 'Superman is Clark Kent'. In any case, no one should deny that true identity statements can be informative in a way that a priori knowable statements can not. So, we can restate the argument to show that 'Superman' and 'Clark Kent' must differ in propositional content in contexts involving a priori knowledge. Whence, Strict Referentialism is false.⁴¹

⁴¹ This argument clearly applies to ordinary proper names which are paradigm purely Referential terms. The question is whether it applies to natural kind terms. In Sec. 3.1, fn. 1, I said that Kripke took natural kind terms to function in much the same way as proper names. This led Tichy [1981] to argue, on similar grounds, that the sentence 'water is H₂O', which states a true theoretical identification, should equally be a priori

What the Referentialist must provide us with, if he is not to deny that propositional content is compositional or truth-conditional, is thus an account of why (PS) fails in some but not all epistemic contexts. It is instructive to make a comparison. In *Naming and Necessity* Kripke advanced a cluster of arguments in support of Referentialism, but he never outright endorsed Strict Referentialism.⁴² He admitted that were Strict Referentialism true, (PS) would have to hold in all intentional contexts, but due to the kind of problem just rehearsed, he was reluctant to embrace the antecedent. Kripke also famously argued that true identity statements between e.g. distinct proper names were necessary a posteriori.⁴³ But as my argument makes clear, were he to accept the unrestricted use of (PS), necessarily true identity statements would come out a priori. So, the Referentialist owes us an account of why (PS) fails in contexts involving a priori knowledge such that we can block the argument and allow for the possibility of a posteriori necessities. I shall not go into any details here as to how this might go. Kripke's own idea was that the mode of fixing the reference of a singular term could somehow enter into the propositional content of a sentence containing an epistemic operator. What is important for present purposes is that Strict Referentialism is untenable: even the Referentialist must allow for the possibility that descriptive information is truth-conditionally relevant.⁴⁴

knowable. His argument rests, however, on the assumption - not spoken to by Kripke - that chemical natural kind terms like 'H₂O' are also purely Referential. If 'H₂O' is short for a *de facto* rigid definite description like 'the stuff that for every oxygen-atom has to hydrogen-atoms', then the argument is blocked since (PS) fails to apply. But the argument would presumably hold for identities between manifest natural kind terms like 'furze' and 'gorse'; cf. also Kripke [1994, p. 378].

⁴² Kripke's reservations can be found in [1980, pp. 20-1, 127-8, 134-5] and [1994, p. 353, fn. 10].

⁴³ Cf. [1980, pp. 100-5]. The necessity of the statement $a = b$ stems from the fact that all purely Referential terms a and b are rigid designators and true identities between non-empty rigid designators are necessary: $a = b \rightarrow \Box a = b$.

⁴⁴ Note that it is potentially unstable to endorse the applicability of (PS) in all belief contexts but in no knowledge contexts. Suppose S knows that a is F . (i) If S knows that a is F , then S believes that a is F . We tend to think that belief is necessary for knowledge. But if S believes that a is F , then, by (PS), S also believes

If, on the other hand, we assume that singular terms have descriptive content as spelled out by Descriptivism, then we can deny the step from (33) to (34) on the grounds that applicability of (PS) in intentional contexts requires preservation of descriptive content. The fact that Lois Lane associates different descriptive properties with 'Superman' and 'Clark Kent' explains the falsity of (34): she does not believe, hence does not know, that the masked super-hero is the shy journalist. Or consider the sentence:

(35) Water is H_2O ,

and suppose again that W_1 is Earth and W_2 is Twin Earth. Suppose also that ' H_2O ' is shorthand for the *de facto* rigid definite description 'the stuff that for every oxygen-atom has two hydrogen-atoms'.⁴⁵ Then we can represent (35) as follows:

fig. 7	W_1	W_2
	True	True
	False	False

As is familiar from Kripke [1980], (35) is an example of an a posteriori necessity. At some point in history, physical science discovered the micro-structure of water, but what it

that b is F. (ii) If S knows that a is F, then a is F. Knowledge is factive. But if $a = b$, then b is F if a is F. No one disputes that (PS) applies in purely extensional contexts: if $a = b$, then 'a is F' is true iff 'b is F' is true. (iii) If S has a warrant for believing that a is F, then S also has a warrant for believing that b is F. Maybe S has a kind of internalist justification for believing that a is F that does not carry over to her belief that b is F, but it seems that whatever externalist warrant S has for the former will also be a warrant for the latter. Thus if S's belief that a is F is counterfactually sensitive in the sense that S believes that a is F in the closest W_p in which a is F, and does not believe that a is F in the closest W_p in which a is not F, then so will be S's belief that b is F, if $a = b$. If (PS) really applies throughout all belief contexts, then it is hard to see how there could be any near-by W_p in which the one belief had a sensitivity the other lacked. Whenever S believes that a is F, S also believes that b is F, and whenever S fails to believe that a is F, S also fails to believe that b is F. So, if S knows that a is F, then S has a warranted true belief that b is F; so S knows that b is F.

⁴⁵ We moreover suppose that ' H_2O ' is deeply rigid (cf. fn. 20, 29): no matter which W_p we consider as actual, ' H_2O ' refers to H_2O . Maybe it is conceivable, or even physically possible, that different 'kinds' of sub-atomic stuff can have two hydrogen-atoms for every oxygen-atom, but it seems to me that they will all be H_2O .

discovered was a necessity: there are no counterfactual W_P in which water is not H_2O . Both 'water' and ' H_2O ' are rigid designators, and true identities between rigid designators are necessary. The question is how to explain this. First of all, what can the Referentialist say? On his account, (35) expresses the singular proposition $\langle H_2O; H_2O; \text{the identity relation} \rangle$ provided we take W_1 as actual, i.e. as our context of acquisition. This proposition is necessary since there is no W_P in which H_2O is not H_2O , and is displayed in fig. 7 by the fact that the first horizontal proposition consists of nothing but truths. In our notation, the sentence 'water is H_2O ' is true-in W_1 , and is, in W_1 , true-of W_2 . But how can (35) come out a posteriori on the Referentialist's account? An a posteriori knowable proposition is such that one can fully understand it without knowing that it is true. So, it must be possible to fully grasp $\langle H_2O; H_2O; \text{the identity relation} \rangle$ without knowing that it is true. But how can that be if, as the Referentialist has it, 'water' lacks associated descriptive content? Presumably, the Referentialist will invoke semantically irrelevant modes of presentation: if one is presented with water, not as H_2O , but as the watery stuff, then one can understand (35) without knowing its truth-value. However, the worry lingers on. Compare with:

(36) Water is water,

which in W_1 expresses the *same* necessarily true singular proposition and yet is clearly a priori knowable. Here the Referentialist must say that it is not possible to fully grasp $\langle H_2O; H_2O; \text{the identity-relation} \rangle$ without knowing that it is true. But why the difference in epistemic access to the very same proposition when (35) and (36) have no associated descriptive content?

In any case, we can offer a satisfactory Two-Dimensional analysis if we avail ourselves of descriptive content. What someone knows when she understands (35) is the descriptive proposition that the watery stuff is H_2O , and one can clearly understand this proposition without knowing that it is true. So, in order to know the singular proposition expressed by (35) in W_1 , one must know which kind of stuff is the watery stuff in W_1 , and that takes empirical investigation. What one knows when one understands (35) is the

diagonal proposition but, as pictured in fig. 7, this proposition consists of both truths and falsehoods. So, in order to know whether (35) is true, one must know which W_P plays the role as actual, i.e. whether one is in W_1 where the watery stuff is H_2O or in W_2 where the watery stuff is XYZ, and obviously one can only know that a posteriori. Both (35) and (36) expresses in W_1 the same necessarily true singular proposition. The difference lies in our epistemic access to that proposition due to the different descriptive contents that the two sentences have associated. Someone who grasps the descriptive proposition associated with (36) knows that it expresses a necessarily true singular proposition no matter which W_P is considered as actual. She therefore knows a priori that (36) is necessarily true. But in the case of (35), grasp of the descriptive content falls short of knowing its truth-value since it is dependent on which W_P is actual.⁴⁶

How is the a posteriori necessity of (35) accounted for by the Descriptivist when no appeal is made to singular content? Well, what someone knows when she understands (35) is the *rigidified* descriptive proposition that the actual watery stuff is H_2O . Obviously, someone can fully grasp this proposition without knowing that it is true: being H_2O was precisely not one of the watery properties. In order to know the truth-value of (35), one must know which kind of stuff is the watery stuff in W_A , but one can know that only after having carried out a substantial empirical enquiry. Nevertheless, when the sentence 'the actual watery stuff is H_2O ' is uttered in W_1 , the proposition it is associated with is necessarily true. A true identity between two rigid designators is necessary. So, it would seem that the very same rigidified descriptive proposition can be both necessary yet only a posteriori knowable as true. In other words, we arrive at theoretical identifications in two stages. First of all, by hypothesising about how to describe various possible cases, we come to know a priori that (if water exists, then) water is the actual watery stuff of our

⁴⁶ Lewis [1999b, pp. 296-7] and Jackson [1998a, pp. 84-6] both think it is distinctive of Two-Dimensionalism that no proposition is both necessary and a posteriori; what qualifies as an a posteriori necessity is the interpreted *sentence* (35).

acquaintance. This involved three kinds of thought experiments about what to say when Twin Earth is taken as: a candidate for W_A , a remote place in W_A , a counterfactual W_P . Secondly, by empirical discovery, we come to know a posteriori that H_2O is the watery stuff of our acquaintance in W_A . Given these two pieces of knowledge, it is not a further empirical question whether water is H_2O . We can simply infer by transitivity of identity, and hence come to know the a posteriori necessity that water is H_2O .⁴⁷

Let me sum up. I have argued that we should settle for Causal Rigidified Descriptivism according to which a term like 'water' is best seen as short for 'the actual watery stuff of our acquaintance'. This view handles the Modal Argument, it accounts for Twin Earth in its various guises, and it blocks the Intentional Argument. Understanding sentences containing 'water' is to know the conditions under which such sentences are true-in W_P , and to know the conditions under which they are true-of W_P . Grasp of descriptive truth-conditions suffices to know W_A -truth-conditions, and grasp of rigidified descriptive truth-conditions is sufficient to know W_P -truth-conditions. The Two-Dimensional analysis highlights that knowledge of - rigidified - descriptive content is a priori, whereas knowledge of singular content is a posteriori. Descriptive content yields truth-conditions for all W_P qua actual, and so is knowable independently of which W_P is in fact actual. This is what our diagonal proposition records: 'water is watery' is true-in W_P iff the watery stuff in W_P is watery in W_P . Moreover, rigidified descriptive content yields truth-conditions for all counterfactual W_P independently of which W_P is actual: 'water is watery' is true-of W_P

⁴⁷ In general, anything of the form @s will be necessary a posteriori. Take s to be 'the watery stuff is H_2O '. @s thus entails @s, since on Earth 'the actual watery stuff is H_2O ' is true-of every W_P . But @s does not entail $\mathcal{J} @(@s)$ since 'the actual watery stuff is H_2O ' is not true-in every W_P ; it is false-in Twin Earth. In Evans' terminology, 'the actual watery stuff is H_2O ' is a superficial necessity, but a deep contingency. These two notions of necessity accommodate Kripke's argument in [1980] that it is metaphysically impossible, hence inconceivable, that water is not H_2O if water is H_2O . What is conceivable, hence epistemically possible, is that an epistemic counterpart of water - a watery stuff - is not H_2O . But this is just the distinction between the necessity of what 'water is H_2O ' actually expresses and what 'water is H_2O ' would express had another W_P been actual; cf. also Yablo [2000].

iff the watery stuff in W_A is watery in W_P . But one cannot know the singular content of 'water' a priori since it is dependent on which W_P is in fact actual. If we take Earth as actual, then the singular content of 'water' is $\langle H_2O \rangle$, but if we take Twin Earth as actual, then the singular content of 'water' is $\langle XYZ \rangle$. This is what our horizontal propositions depict. Only after the empirical facts are in, can we pin down singular truth-conditions for all counterfactual W_P : 'water is watery' is true-of W_P iff H_2O is watery in W_P . This difference in epistemic properties between descriptive and singular content will prove important when we now return to our problem about Incompatibilism from Chapter 1.

Chapter 5. Beyond Incompatibilism

5.1. Three Kinds of External Dependence

We have seen in Sec. 4.4 that an utterance of a simple sentence 'N is F' can be assigned two distinct sets of truth-conditions corresponding to the two ways 'N is F' can be true at a W_P . If we think of W_P as actual, i.e. as a possible context of utterance or acquisition, we can give W_A -truth-conditions, and if we think of W_P as counterfactual, i.e. as a circumstance of evaluation, we can give W_P -truth-conditions. According to Descriptivism, what it standardly takes for a competent speaker S to understand an utterance of 'N is F' is knowledge of W_A -truth-conditions, and to know them is to know:

Descriptive truth-conditions: 'N is F' is true-in W_P iff ϕ in W_P is F in W_P ,

which gives the conditions under which 'N is F' is true at all W_P considered as actual. S's grasp of descriptive truth-conditions gives her a fairly comprehensive mastery of the assertoric use of 'N is F', but if she does not know how 'N is F' embeds inside the scope of modal operators, her understanding is essentially incomplete. What S must also know is that N is a rigid designator, i.e. that ϕ must be rigidified such that the referent of N at a counterfactual W_P is whoever satisfies ϕ in W_A (whichever W_P we consider as actual). S's understanding must also encompass knowledge of:

Rigidified Descriptive truth-conditions: 'N is F' is true-of W_P iff @ ϕ is F in W_P ,

which ensures knowledge of a set of conditionals: if o satisfies ϕ in W_A , then 'N is F' is true-of W_P iff o is F in W_P ; if o* satisfies ϕ in W_A , then 'N is F' is true-of W_P iff o* is F in W_P ; etc. To know the rigidified descriptive truth-conditions is to know W_P -truth-conditions, and is, following Descriptivism, all understanding takes. According to Referentialism, however, understanding 'a is F' requires knowledge of:

Singular truth-conditions: 'N is F' is true at W_P iff o is F in W_P ,

which gives the conditions under which 'N is F' is true-in W_A and true-of W_P , i.e. at various counterfactual W_P . On this account, there is no significant distinction between knowledge of W_A -truth-conditions and knowledge of W_P -truth-conditions; in both cases, understanding requires knowledge of singular content. Someone who knows the singular truth-conditions has relevant empirical information such that she is able to detach the consequents of our conditionals. She knows which object satisfies ϕ in W_A , and so knows which object N refers to. According to Referentialism, understanding requires knowledge of reference. But this is exactly what Descriptivism disputes. A speaker who knows which object N picks out in W_A is able to advance from knowledge of rigidified descriptive truth-conditions to knowledge of singular truth-conditions, but this is not a speaker who has a better understanding than someone who does not know the reference of N. This is a speaker who knows more truths, and not a speaker who has a better grasp of truth-conditions. To know the reference of N is an a posteriori matter which at least in a wide range of cases is an over-demanding constraint on competence with N to impose on speakers.

Bear in mind the different ways in which descriptive and singular content is individuated on the two views. We have seen in Sec. 3.2 that according to Descriptivism, propositional content is individuated *intentionally*:

(DCA) If 'P' and 'Q' have the same propositional content, then a fully competent speaker S believes that P iff S believes that Q.

On this view, propositional content is essentially cognitive in nature such that singular terms have the same *descriptive content* iff they are intersubstitutable in *intentional* contexts *salva veritate*. Take 'water' and 'the watery stuff of our acquaintance' which arguably have the same descriptive content: S believes that water is wet iff S believes that the watery stuff of her acquaintance is wet. After all 'water is wet' and 'the watery stuff of

our acquaintance is wet' have the same W_A -truth-conditions, and so it would seem that S believes the same thing about how W_A is. Or consider 'Julius' and 'the inventor of the zip'. Someone who is competent with 'Julius' believes that Julius was ingenious iff she believes that the zip-inventor was ingenious. So, it would seem that *singular terms coincide in descriptive content iff they are intersubstitutable in intentional contexts salva veritate*. There are, however, two difficulties.

The first arises in *mixed contexts*. Suppose S sincerely and on reflection assents to 'water might not have been watery'. It is true that water might not have been watery since there is a W_P in which H_2O is not watery. By Disquotation (D) we infer that S believes that water might not have been watery. But substituting 'water' for 'the watery stuff' fails to preserve truth-value. It is false that the watery stuff might not have been watery since there is no W_P in which the watery stuff is not watery. One way to remedy such troublesome cases is to place a constraint on the belief ascribing sentences such that singular terms coincide in descriptive content iff they are intersubstitutable in *non-modal* intentional contexts *salva veritate*. Another way is to insist on rigidification such that the descriptive content of 'water' is given by 'the actual watery stuff'. That way, supplanting one for the other in intentional contexts will not issue in different truth-values even if the embedded sentences contain modal vocabulary. The point is that if S fully understands 'water might not have been watery', then she will know that a rigid designator cannot be substituted for a flexible designator in modal contexts without change in truth-value.

The second difficulty arises in cases where S sincerely and on reflection assents to a sentence like 'water is XYZ'. The question is whether we can reasonably use (D) to infer that S believes that water is XYZ. We are inclined to think that although our beliefs are often mistaken, nevertheless, they could have been right. To believe something is to represent a way things might be and to believe that things are that way. But what is it S purports to represent when S assents to 'water is XYZ'? If there is no W_P in which water is XYZ, then what is the possible object of belief that 'water is XYZ' refers to - the empty

proposition? The worry is that if we ascribe to S the belief that water is XYZ, then S would seem to believe a necessary falsehood, but if we do not, then we are unfaithful to S's linguistic behaviour. That is, we should aim to ascribe beliefs such as to make most sense of our fellow speakers without having an error-theory of their assertions. I suspect there are two plausible things S might believe. We rightly describe S as believing that water is XYZ, since what this means, according to Descriptivism, is that S believes that the watery stuff of her acquaintance is XYZ. Although this ascription is false, it could have been true. The watery stuff around here is H_2O , but it might have been XYZ. Or maybe what S believes is not the necessary falsehood of what 'water is XYZ' expresses on Earth, but the necessary truth of what 'water is XYZ' might have expressed had S been on Twin Earth. That is, S has the false but possibly true belief that S is in a W_P in which 'water is XYZ' expresses a truth.¹

The singular content of a term is, however, clearly not individuated by its behaviour in intentional contexts. Although 'water' and ' H_2O ' coincide in singular content in W_A , we have argued that substituting one for the other can easily result in different belief ascriptions. It is not the case that S believes that water is wet iff S believes that H_2O is wet. Singular content is not individuated intentionally but by the identity conditions of its singular constituents: two singular terms have the same singular content iff their referents are numerically identical. Referentialism is the view that all there is to the content of a

¹ We know that if water is H_2O , then it is necessary that water is H_2O . But surely we think we can conceive of a possible scenario in which water is not H_2O , but say XYZ. There is, as Kripke [1980] argued, an *illusion of contingency* which must be explained away. The foregoing suggests that what is conceivable is that there is a stuff which has all and only the watery properties H_2O actually has but is not H_2O . So, the apparent possibility that water is not H_2O is accounted for by the genuine possibility that the watery stuff is not H_2O . Cf. also Stalnaker [1999a, pp. 123-7] who argues that sometimes what a speaker believes is not the horizontal but the diagonal proposition, e.g. when someone asserts that Hesperus is Mars, we intuitively ascribe to him the belief that the solar system is arranged such that Mars appears in the evening at the same place where Venus in fact appears. That is, sentences like 'Hesperus is Mars' and 'water is XYZ' have 'contingent' diagonal propositions. See also Yablo [2000], and Sec. 5.3.

singular term is its bearer. On this view, propositional content is individuated *modally* such that two singular terms N and ϕ have the same propositional content iff they are intersubstitutable in *modal* contexts *salva veritate*:

(RCA) If N and ϕ have the same propositional content, then ' N is ϕ ' expresses a necessary truth.

Singular content sustains (RCA). If two singular terms coincide in singular content then they refer to the same object at every counterfactual W_P , and so are interchangeable in modal contexts *salva veritate*. In W_A 'water' and ' H_2O ' both express the singular content $\langle H_2O \rangle$. This corresponds to the fact that both are rigid designators with respect to H_2O , hence that 'water is H_2O ' expresses the necessary singular proposition $\langle H_2O; H_2O; \text{identity-relation} \rangle$. Or take 'Superman' and 'Clark Kent'. Both are modally rigid with respect to Superman, hence 'Superman is Clark Kent' expresses the necessary singular proposition $\langle \text{Superman}; \text{Superman}; \text{identity-relation} \rangle$. So, *singular terms coincide in singular content iff they are intersubstitutable in modal contexts salva veritate*.

According to Referentialism, propositional content is thus individuated by its *modal properties* independently of how we conceptualise it: singular content is simply a matter of how the world is regardless of its modes of presentation. E.g. water and H_2O are the same kind of stuff but thought of in different ways. What S believes when S believes that P is essentially a function of how the world is such that if one changes how the world is, one changes the content of S 's belief, and hence the belief state that S is in. Thus the Twin Earth Argument aimed to show that one could change the content of beliefs about the world by changing the way the world is while fixing the way the world appears. According to Descriptivism, however, propositional content is individuated by its *cognitive properties*: descriptive content captures certain modes of presentation of the world and it is possible that different aspects of the world be presented under one and the same mode of presentation. E.g. water and twin-water are different kinds of stuff but are thought of in the

same way. What S believes when S believes that P is essentially a function of how the world appears to S from the inside, as it were, such that if one changes how the world appears, one changes the content of S's belief and hence the belief state that S is in. Thus the Intentional Argument aimed to show that one could change the content of beliefs about the world by changing the way the world appears while fixing the way the world is.

This strongly suggests that we should think of the descriptive aspect of mental content as *narrow* and of the singular aspect of mental content as *wide*. We saw in Sec. 2.2 that Semantic Internalism relied on Descriptivism. Descriptive content is narrow in that it is common to intrinsic duplicates despite their different environmental embeddings. My doppelgänger on Twin Earth and I both believe that the watery stuff is wet since water and twin-water have the watery properties in common. In Sec. 2.3, however, we saw that Semantic Externalism relied on Referentialism. Singular content is wide in that it is not shared amongst intrinsic duplicates. Due to the fact that water and twin-water have distinct micro-structures, my doppelgänger on Twin Earth believes that twin-water is wet, whereas I believe that water is wet. We can thus characterise the difference between descriptive and singular content by saying that the latter is *dependent on the external environment* in a way that the former is not. So far this relation of dependence has been spelled out in terms of *object-dependence* in the sense that had the referent of a singular term N not existed, the content of N would not have existed. Singular content is object-dependent: my tokens of 'water' express the singular content $\langle \text{H}_2\text{O} \rangle$, but had water not existed, they could not have expressed just that content. Descriptive content, on the other hand, is object-independent: my tokens of 'water' express the descriptive content $\langle \text{the watery stuff} \rangle$ whether water exists or not. Object-dependence involves a strong external constraint on concept possession: one can only entertain an object-dependent thought if one, or one's fellow speakers, have had causal encounters with the particular objects on which one's thought is dependent. So, intrinsic duplicates may easily fail to share object-dependent thoughts. Now, as we saw in Sec. 1.6, there is clearly space for a weaker relation of external dependence. That is, our

definition of wide content as content intrinsic duplicates may fail to have in common does not require that wide content be object-dependent. Our problem was how content could be wide, but not object-dependent, if Referentialism were true. Considerations about Dry Earth cases seemed to show that there could be no merely weak form of Semantic Externalism. But now that we have reason not to hold out much hope for Referentialism, the question arises whether there could be a notion of object-independent wide content. I think so. We have learned from Twin Earth that the descriptive content of 'water' must be causally constrained: <the watery stuff of our acquaintance>. One of the cluster properties associated with 'water', as used by my speech community, is the property of being the kind of stuff that *I*, or *my* fellow speakers, causally interact with. Now take my doppelgänger on a remote Twin Earth in W_A . He has also learned from thought experiments that descriptive content must be causally constrained, but he does not associate with 'water', as used in his speech community, the property of being the kind of stuff that *I*, or *my* fellow speakers, causally interact with. What he associates with 'water' is the property of being the kind of stuff that *he*, or *his* fellow speakers, causally interact with. This is precisely why he cannot have water beliefs: he is not acquainted with the watery stuff I am acquainted with, and why I cannot have twin-water beliefs: I am not acquainted with the watery stuff he is acquainted with. In other words, my utterance of 'water is wet' is true iff the watery stuff of *my acquaintance* is wet, whereas his utterance of 'water is wet' is true iff the watery stuff of his acquaintance is wet. So, causally constrained descriptive content is wide in the sense that it is not invariably shared by intrinsic duplicates within W_A . To have a belief with such content is to have a mental property that fails to supervene on purely intrinsic properties. We shall call such content *context-dependent* in that it varies with shifts in contexts of acquisition - Earth or Twin Earth - within W_A . Crucially, causally constrained descriptive content is not object-dependent: had I been on Dry Earth, I could still have believed that the watery stuff of my acquaintance is wet; it is just that nothing would make that belief true. So, object-dependence entails context-dependence - my doppelgänger on a remote Twin

Earth in W_A cannot believe that water is wet - but context-dependence does not entail object-dependence. Once Semantic Externalism is separated from Referentialism, there is hope for a notion of object-independent wide content.

Note finally that purely descriptive content is both object- and context-independent. If I believe that the watery stuff is wet, then so does my doppelgänger on Twin Earth, and so would I had I been on Dry Earth. To have a belief with purely descriptive content is to have a mental property that supervenes on intrinsic properties. But, as I shall now argue, although such content is shared by intrinsic duplicates within W_A , it is not necessarily shared by intrinsic duplicates across all W_P . Even purely descriptive content is *world-dependent*. So, we arrive at a threefold distinction between object-, context- and world-dependent content. What I shall do in the following is to mark out these distinctions, and in particular buttress the claim that some mental content is narrow. I shall first say something about what narrow content could plausibly consist in and then discuss some objections that have been advanced against narrow content.

5.2. What is Narrow Content?

We have stipulated narrow content to be a mental property that supervenes on intrinsic, physical properties of individuals; i.e. a property necessarily shared by doppelgängers such that, in every W_P , two intrinsic physical duplicates are narrow content duplicates.² In order to get clear on what kind of *supervenience* is in play, consider three examples of kinds of properties:

(i) The property of being *square* is a property that supervenes on internal structure. Necessarily, an object is square iff its surface is shaped by four equally long straight lines at

² As I said in Sec. 1.1, fn. 1, there is an issue about which properties constitute the supervenience base. It is standard to include physical, and at times functional, properties. Nothing I say should hang on whether other kinds of properties can go into the base. The important point is that my doppelgänger is someone who is just like me from the skin in.

right angles. Thus if an object is square then, necessarily, any internal duplicate of it will also be square. No change in relational properties can make it true that a doppelgänger of a square object is not square. Moreover, for an object to be square just *is* for it to have a certain set of intrinsic, geometrical properties - having a shape with four equally long, straight lines and right angles - and to understand what it is for an object to be square is just to know that it has all these properties.

(ii) The property of being *water-soluble* is also a property that supervenes on internal structure. Necessarily, an object is soluble iff under *normal conditions* it would dissolve if immersed in water. There is some internal physical state of the object which is causally responsible for the behaviour such that any internal duplicate of it is *normally* itself water-soluble. Hence, for an object to be water-soluble is not just to be in a certain internal state in the sense that if all you know is confined to that internal state, you do not know whether it is water-soluble; you have to know something about the causal connections between being in that state and behaving in a certain way on being put in water.

(iii) The property of being a *footprint* is a property that fails to supervene on internal structure. Necessarily, something is a footprint iff it has a *normally* foot-shaped imprint and is *caused by a normal foot*. An intrinsic duplicate of a footprint is not a footprint if not caused by a normal foot. A foot-shaped imprint caused, say, by the way the waves happen to fall on the sand is not a footprint. Again, if all you know is confined to the intrinsic properties of a particular imprint, you do not know whether it is a footprint; you have to know something about the causal connections between being a footprint and being caused by the impact of a foot.³

³ The examples are modelled on Jackson & Pettit [1993, pp. 271-2] and Stalnaker [1999a, pp. 171-4]. See also Davidson [1996, pp. 335-8] who argues that the property of being *sunburned* presupposes certain causal relations between the sun and the sunburned skin. Skin may be qualitatively indistinguishable from sunburned skin and yet not sunburned if it does not have the right causal history. Such skin may share a narrow condition of sunburned skin: sunnishburn is just like sunburn except no particular causal connections need be sustained.

With this background, we may say that being square *globally supervenes* on internal structure. Global supervenience is a relation between W_P such that if two W_P are exactly alike in which objects are shaped by four equally long, straight lines at right angles, then they are exactly alike in which objects are square; regardless of how they may otherwise differ. Alternatively, we can put roughly the same claim *locally* as a relation between individuals in W_P :

S strongly supervenes on B iff for any individuals x and y and for any W_P and W_P^* , if x in W_P is B-indiscernible from y in W_P^* , then x in W_P is S-indiscernible from y in W_P^* ;

where the supervenience-property S is being square and the base-property B is having certain geometrical properties.⁴ That is, the property of being square is *cross-world narrow*, where a property P of x is cross-world narrow iff *in every* W_P any doppelgänger of x has P. Being water-soluble, however, does not strongly supervene on internal structure. An object may be intrinsically similar to a soluble object and yet not dissolve in water if some anomalous environmental conditions obtain. We can imagine a W_P with relevantly deviant laws of nature in which a doppelgänger of a soluble object in W_A does not dissolve if put in water. What is needed is:

S weakly supervenes on B iff for any individuals x and y and for any W_P , if x and y are B-indiscernible in W_P , then x and y are S-indiscernible in W_P ;

where S is being water-soluble and B is a certain internal physical structure. That is, the property of being water-soluble is *intra-world narrow*, where a property P of x is intra-world narrow iff *in every* W_P where x has P any doppelgänger of x has P. Within W_P internal duplicates are duplicates with respect to whether or not they are soluble, and so to find an object which is internally exactly like a water-soluble object without being water-

⁴ Cf. Kim [1994, pp. 577-9].

soluble, you must go to a different W_P . Intra-world narrow properties are only shared by doppelgänger across nomologically identical W_P , i.e. W_P consistent with the actual laws of nature. Lastly, being a footprint is not even intra-world narrow since there are W_P *within which* a doppelgänger of a footprint fails to be a footprint. Does this make being a footprint a *pure* relational property? No. Not everything caused by the impression of a foot is a footprint. What makes an imprint in the sand a footprint is not just its aetiology, but also its foot-shapeness, i.e. a certain internal distribution of grains of sand in the imprint. It follows that from the wide property of being a footprint, we can *factor out* the narrow property of being a *foot-shaped imprint*. This property is intra-world narrow: a doppelgänger in a W_P of a foot-shaped imprint x is itself a foot-shaped imprint *only if x is also a foot-shaped imprint in W_P* , despite any difference in causal histories. Where footprints owe their identity to *particular* causes, foot-shaped imprints are independent of any such. Something is a foot-print only if caused by a foot; something is a foot-shaped imprint if caused by a foot, the waves or what not. An imprint which is intrinsically indistinguishable from a foot-print is a foot-shaped imprint, but is a footprint only if caused by a foot. The property of being a foot-shaped imprint is, however, still dependent on *general* facts extrinsic to the sand. In a counterfactual W_P where feet have abnormal shapes, hence where the standard for being a foot is different, the imprint in the sand is not even a foot-shaped imprint. That is why we need the condition that x be a foot-shaped imprint in W_P . To form the narrow property out of the property of being a footprint, we must qualify the imprints to include only the *normal* shapes that feet make under *normal* conditions (and thus assume that human footprints have common characteristics). So, the property of being a foot-shaped imprint is not cross-world narrow.⁵

⁵ This suggests that it is *physically* necessary that something is a footprint iff it has a normally foot-shaped imprint and is caused by a normal foot. If, however, we rigidify on 'normally'/'normal', then there is a case for saying that the bi-conditional is also *conceptually* necessary since it is arguably what someone knows who understands what 'footprint' means. The same goes *mutatis mutandis* for water-solubility.

Now, which of these are mental properties akin to? Distinguish again between the content of an attitude and the property of having an attitude with that content. It is clear that the former is not an intrinsic property. The representational character of content tells us that content is, at least partially, individuated by its truth-conditions, but truth-conditions specify how the *world* must be in order for what is thought or said to be true or false. The truth-conditions of most of the declarative sentences in our language can only be stated in terms not internal to any particular individual. But if *content is not intrinsic* in the sense that it can be understood in purely internal terms as the property of being square can, then neither is the property of having an attitude with that content. That is, if mental properties were like being square, then not only would they supervene on an individual's intrinsic properties, but it would also be possible to determine which thoughts an individual is apprehending by investigating only what is in her head, not her language community nor her physical surroundings. Someone who knew all of an individual's intrinsic properties would thereby also know the individual's mental properties. But mental properties are not wholly internal in this sense. Content-bearing mental properties are no doubt states located within individuals, but they can only be understood relationally, i.e. by describing how we actually or counterfactually interact with external objects.⁶ The point is that although mental properties cannot be explained entirely in terms internal to an individual, they can nevertheless *supervene* on an individual's intrinsic properties - if carefully cashed out. It is clear that the Semantic Internalist should not model supposedly narrow content on properties like being square. Unlike being square, narrow content does not strongly supervene on internal structure, but is a function of internal structure plus laws of nature and linguistic practice. In W_P with abnormal laws of nature or deviant linguistic practices, there are doppelgängers of us that do not share our narrow contents. Narrow content is not

⁶ Similarly, a footprint is a property of the sand, but one can only understand what a footprint is if one knows how it is typically caused. E.g. the traces made in the sand by a crawling ant of a recognisable caricature of Churchill do not *depict* Churchill. In Putnam's [1981, p. 5] words. "Thought words and mental pictures do not *intrinsically* represent what they are about." Cf. also Jackson & Pettit [1993, pp. 271-2].

independent of *any* causal interactions with the environment, like squareness is, but only of which interactions are actual and which possible. Narrow content is more like being water-soluble in that the intrinsic properties, shared by doppelgängers, govern interactions with actual and possible environments. A sugar-cube that never actually gets dissolved, because never placed in water, does not thereby lose its solubility as long as it would dissolve were it placed in water.

Take a mental property like having the thought that the stuff that goes under the name 'water' is wet. My intrinsic properties govern which actual and possible state of affairs makes the thought true. On Earth H_2O makes the thought true, but had I been on a Twin Earth in W_A , XYZ would have made the thought true. Having a narrow content thought is thus independent of which *particular* environment one inhabits, but governs what one's thought would be about given actual and possible environments. Contrast with the property of having the thought that water is wet. Unlike a narrow content thought, this wide content thought has a *particular* external dependency. On Earth where I interact with water, I can entertain this thought, but had I been on Twin Earth I would not have interacted with water, and so could not have entertained this thought. What narrow content thoughts have is a *general* external dependency in that it is not shared by doppelgängers across all W_P . Just as we can imagine a doppelgänger in a far-fetched W_P of a sugar-cube in W_A which fails to dissolve if put in water due to peculiar physical laws in W_P , we can imagine a doppelgänger of mine in a far-fetched W_P , who does not share my narrow contents, if in W_P the language is very different or the laws of nature are different in a way that affects the causal roles of those of my brain-states upon which my narrow content beliefs depend. So, narrow content mental states are *context-independent* properties of individuals in that they are necessarily shared by doppelgängers in *different* contexts of

acquisition within the *same* W_P ; such states are not *world-independent* as they are not necessarily shared by doppelgängers across W_P .⁷

The mental property of having the thought that water is wet is more akin to the property of being a footprint. In light of the Twin Earth Argument, we must say that since at least part of the content of that thought is wide, i.e. not even intra-world narrow, the property of having a thought with that content does not supervene on the individual's intrinsic properties. Earthlings can think that water is thirst-quenching, but their doppelgängers on Twin Earth cannot since they do not sustain the requisite causal connections that their Earthly doppelgängers do. Nevertheless, the *abstraction-process* that we used to factor out the narrow property of being a foot-shaped imprint from the wide property of being a foot-print, can also be deployed to factor out the narrow content of the wide content thought that water is wet. One starts off with two intra-world doppelgängers who differ in wide content. Then one abstracts from their particular context-dependency until one finds out which mental properties they have in common. Since these will weakly supervene on their intrinsic properties, they will be narrow. *Narrow content is what remains constant under counterfactual variation in wide content.*⁸ On Earth I believe that water is wet, but had I been on Twin Earth, I would have believed that twin-water is wet. On Twin Earth my doppelgänger believes that twin-water is wet, but had he been on Earth,

⁷ The foregoing owes much to Jackson & Pettit [1993, pp. 271-3] and Lewis [1999b, p. 315] who have argued that some content is in this sense intra-world narrow. Note a special difficulty about *egocentric* content. Suppose my worldmate doppelgänger and I both utter 'I have a sore tooth'. Due to the mere fact that we are distinct individuals, our respective utterances have different truth-conditions. So, given our definition of narrow content, the content of my belief that I have a sore tooth is wide, but wide content was supposed to be individuated by reference to *external* circumstances. Maybe, as Lewis [1979] has suggested, such *De Se* attitudes are self-ascriptions of properties, e.g. my doppelgänger and I have in common the property of having a sore tooth. See also Braddon-Mitchell & Jackson [1996, pp. 216-9], and Sec. 5.6.

⁸ Strictly speaking, all the abstraction-process warrants is the claim that intra-world doppelgängers share certain mental properties, namely whatever properties map contexts of acquisition onto wide contents. I shall argue in Sec. 5.3 that those common properties are indeed truth-conditional, content-bearing mental states; cf. also fn. 10.

he would have believed that water is wet. That is, our shared internal nature governs the same interactions with actual and possible environments. We both associate an acquaintance condition with 'water' such that if we interact with different kinds of stuff, we believe different things. But we can factor out a narrow condition, namely the property of being acquainted with a watery stuff. Narrow content is the part of our belief contents that stays fixed in such counterfactual circumstances.

Let me finally say a few words about the motivation behind the invocation of narrow content. So far we have focused on the representational properties of the contents of our attitudes, but it is clear that those contents also have causal-explanatory properties: we tend to behave in ways that satisfy our desires were our beliefs true. Consider an example. My doppelgänger and I are both thirsty. I am confronted with a glass₁ of water and my doppelgänger with a different but subjectively indistinguishable glass₂ of twin-water. We both utter the sentence 'that glass contains water' and reach out to satisfy our thirst. What are the conditions under which our respective utterances are true? Again, we are faced with a choice. If we go for *singular truth-conditions*, then my utterance is true iff glass₁ contains water, whereas my doppelgänger's utterance is true iff glass₂ contains twin-water. Consequently, I express my belief that glass₁ contains water, whereas my doppelgänger expresses his belief that glass₂ contains twin-water. But if we go for *purely descriptive truth-conditions*, then both our utterances are true iff the demonstratively identified glass having descriptive properties F, G, etc. contains a watery stuff. In such a case where distinct objects have superficially indistinguishable appearances, it is natural to think that grasp of the descriptive, narrow content is sufficient to understand our utterances; it is certainly not mandated that one be able to distinguish glass₁ from glass₂ or water from twin-water. Why do we think that? Well, one consideration concerns *phenomenology*. My doppelgänger and I share a certain psychological perspective: from an internal point of view, the world seems the same to us. There is a glass₁ in front of me containing water, but for all I could tell, there might have been a glass₂ containing twin-water in front of me.

Likewise for my doppelgänger. He is confronted with glass₂ containing twin-water, but for all he knows, he might have been confronted with glass₁ containing water. Narrow content is what captures this common way things seem to us. But why is this qualitative aspect important? Because it seems to suffice for psychological explanation. I reach out for glass₁ to quench my thirst, and my doppelgänger reaches out for glass₂ to quench his thirst. Had I been confronted with glass₂ containing twin-water, I would have reached out for glass₂, and had he been confronted with glass₁ containing water, he would have reached out for glass₁. Same narrow content belief, but different wide content beliefs, implies same behaviour. The difference in wide content is explanatorily redundant.⁹ Or let us modify the example such that I am confronted with a glass of water, and we both know that only my (approximate) doppelgänger is thirsty. He has a belief he would express by saying 'I am thirsty', whereas I have a belief I would express by saying 'he is thirsty'. Our beliefs have the same singular, but different descriptive, contents. Yet we behave in different ways: I reach out for the glass in order to pass it on to my doppelgänger who drinks the water. So, different narrow content beliefs, but same wide content beliefs, implies different behaviour. So, wide content beliefs seem too coarse-grained for all psychological explanation.

⁹ Friends of wide content object that folk-psychological explanations are concerned with behaviour under relational, and not physical, descriptions, and that behaviour thus understood can only be explained by wide content mental states. Thus I reach out for glass₁ since I believe that glass₁ contains water, whereas my doppelgänger reaches out for glass₂ since he believes that glass₂ contains twin-water. The question is thus how fine-grained behaviour should be individuated. Our behaviour can be given the same non-relational descriptions and that may suffice for the purposes of psychological explanation. To go into details here would take me too far a field. The crucial point is that what the narrow content theorist should claim is not that wide content beliefs are explanatorily idle, but merely that narrow content beliefs play an irreducible role in explanation of ordinary belief-desire psychology.

5.3. Some Objections against Narrow Content.

So much by way of characterisation and motivation of narrow content. It is time to face some objections. Despite its intuitive appeal, many philosophers have found the notion of narrow content hard to make sense of. The scepticism has been led by Block [1991, 1999] and Stalnaker [1999a]. In order to get clear on the dialectic of their main point, it is helpful to ask the question: is the distinction between narrow and wide content a distinction *between predicates* such that some predicates are narrow and some wide or is it a distinction *within predicates* such that every predicate has a narrow and a wide component? The problem is this. We started off assuming that 'being water' should be analysed as the narrow predicate 'being the watery stuff'. The Twin Earth Argument proved us wrong. 'Being water' is not even intra-world narrow since doppelgängers within the same W_p fail to share water-beliefs. The response was to add a causality constraint to the cluster of watery properties. That way, my tokens of 'water' do not pick out XYZ on a Putnam-type Twin Earth since XYZ is not the watery stuff of my (or our) acquaintance. So, Twin Earth taught us that 'being water' is a predicate that expresses the property of being the watery stuff of *my acquaintance*. On Twin Earth, my intra-world doppelgänger will also associate with 'water' the property expressed by 'the watery stuff of our acquaintance', but given that he and I are acquainted with distinct kinds of stuff, H_2O and XYZ, his tokens of 'water' will express the property of being the watery stuff of *his acquaintance*. My doppelgänger cannot entertain water-thoughts since he is not acquainted with the watery stuff of my acquaintance, and so 'being water' is a wide predicate. Yet we both associate with 'water' the common property of being a watery stuff, or even the property of being acquainted with a watery stuff. So, we can also learn from Twin Earth that narrow conceptual components can be cut out from 'being water'. But what is to stop the Semantic Externalist from running Twin Earth Arguments on the supposedly narrow, watery predicates. The worry is that if all predicates are susceptible to Twin Earth Arguments, then all one can do by way of

explicating narrow content is allude to functional definitions: narrow content is whatever takes one from a context of acquisition to wide content, e.g. whatever property determines that if I am on Earth, then I have water-thoughts, and if I am on Twin Earth, then I have twin-water thoughts.¹⁰ Presumably, the Semantic Internalist would insist that just as we could factor out the supposedly narrow predicate 'being liquid' from 'being water', so we could make up a narrow predicate 'being liquidish' from 'being liquid'. Let alone the artificiality of such contrived predicates, if any spelling out of the content of 'liquidish' is bound to issue in yet another set of predicates which are equally hostage to a Twin Earth Argument, then Semantic Internalism begins to look futile. Narrow content should not be ineffable! So, to make good the claim that some content is narrow, there ought to be a range of genuinely narrow predicates which fill in the details of 'the watery stuff', and so the scope of the Twin Earth Argument must be limited. Luckily, I think there is. Consider how it goes in Block & Stalnaker [1999, pp. 30-1]:¹¹

"...we are not guaranteed a priori that water is the - or even an - odourless drinkable liquid in rivers and lakes that we have been calling 'water', since it is not even guaranteed a priori that water is a liquid. There is a Twin Earth in which the stuff that they call 'water' is H₂O, as here, but the stuff that they call 'liquid' is virtually all a slippery granular solid [...]. According to this story, water is an exception, one that the residents of Twin Earth would not call 'a liquid' if they knew the scientific facts. (Imagine that on Twin Earth water is rare). Because [...] the counterfactual situation in which this story is true is a twin of ours, an utterance is actually a priori only if the counterpart utterance in that situation is a priori there. Since the counterpart of "water is a liquid" is false there, it cannot be a priori true."

I take it the point here is not that 'water is a liquid' cannot be a priori true *because* there is a counterfactual Twin Earth on which water, i.e. H₂O, is not a liquid. That would be to

¹⁰ For an example of a narrow content theorist who sticks to functional definitions of narrow content, see White [1982] who uses Kaplan's notion of character, and Block [1991] for criticism.

¹¹ For a similar line of thought see White [1982, p. 355], Burge [1982], and Block [1991, p. 50].

confuse a priority with necessity. The point is rather that if 'water is a liquid' were a priori true, then it would express a truth in every W_P qua actual. But an utterance on this Twin Earth of 'water is a liquid' is false, so the sentence cannot be a priori in the sense of expressing a truth in every W_P . This is better but still not good. Why is 'water is a liquid' false on Twin Earth? Because 'liquid' applies to slippery granular solids, and water, i.e. H_2O , is a liquid, and not a solid, on Twin Earth.¹² So, the argument makes the prima facie plausible assumption that 'liquid' functions much like 'water': 'liquid' is a rigid designator of the liquidish stuff. On Earth 'liquid' rigidly picks out liquids like water, but on this Twin Earth 'liquid' rigidly picks out subjectively indistinguishable granular solids. So, just like 'water', 'liquid' presumably has a fixed descriptive content on Earth and Twin Earth, but sentences containing 'liquid' have different truth-conditions. The claim is thus that not only 'water', but also the watery predicates have a *natural kind use*: a predicate like 'being liquid' serves as a rigid designator of *the physical kind* common to the liquidish samples of our acquaintance. Given that 'liquid' refers to whatever physical kind liquids are on Earth, 'liquid' does not refer to solids on Twin Earth even if they are subjectively indistinguishable from liquids.

But why should we accept the analogy between water and liquid? We have previously emphasised that the way the world seems to us, as opposed to the way the world is, plays an important role when we individuate intentional states and explain behaviour. We must thus have at our disposal sufficient linguistic means to facilitate the expression of

¹² Suppose instead that 'water is a liquid' expresses a falsehood on Twin-Earth due to the fact that water, i.e. H_2O , is not a liquid there. Would that imply that 'water is a liquid' is not a priori? Well, we might say different things. If we hold that nothing is water in W_A unless it is a liquid, then there simply is no coherent description of a W_P qua actual in which water is not a liquid, which is not to say that there is no counterfactual W_P in which water is not a liquid. This seems to be very plausible. Or consider the property of being a physical stuff: do we have conceptual space for the possibility that ectoplasm could have turned out to satisfy the watery properties in W_A ? Alternatively, one might respond as we did to the Epistemological Argument, that water is whatever stuff has sufficiently many of the watery properties in W_A , and that perhaps being a liquid is not an indispensable property in this respect.

these, as it were, modes of presentation of the world. One such helpful device would be modally flexible expressions. Let me give an example. Suppose I am in severe pain, and suppose that our concept of pain is the concept of the state that occupies a certain causal role (being caused by tissue damage and resulting in escape behaviour). Science has discovered that what actually fills the pain role (in humans) is the physical state of having C-fibres firing, and so I am in this physical state. Now suppose I were in a *qualitatively identical epistemic situation* in which it is the physical state of having D-fibres firing that fills the pain role (in humans). So, if 'pain' is a rigid designator with respect to the physical 'filler' state, then I am not in pain in this counterfactual situation, and if I am not in pain, then surely I am aware that I am not in pain. In the normal run of things, my pains are transparent to me. But I am not aware I am not in pain in this situation; indeed it seems to me as if I am in pain. So, 'pain' is not modally rigid: what is pain in W_A is not pain in counterfactual W_P .¹³

So, maybe 'liquid' is better modelled on 'pain'. What counts for something to be a liquid is that there is an occupier of the liquid role, and not which physical state that is. So, if, as in the example, liquids fill the liquid role on Earth, then 'liquid' refers to liquids on Earth. But had the liquid role been filled by slippery granular solids, then 'liquid' would refer to those solids; after all the two substances are qualitatively indistinguishable. What the Semantic Externalist must argue, and not just presuppose, is that a predicate like 'being

¹³ Cf. Lewis [1983, 1999b]. Kripke [1980, pp. 147-55] famously denied the premise that the reference of 'pain' is fixed by the *accidental* property of playing the pain role. On his view, the reference of 'pain' is fixed by the *essential* property of being painful, and so 'pain' is a rigid designator of states with this property. Kripke needed rigidity in his argument against physicalism: if pain is C-fiber stimulation, then pain is necessarily C-fiber stimulation. But pain is contingently C-fiber stimulation since any counterfactual epistemic counterpart of pain is pain: "To be in the same epistemic situation that would obtain if one had a pain *is* to have a pain...". So, by *Modus Tollens*, pain is not C-fiber stimulation. I shall not embark upon this argument here. It suffices to point out that neither Lewis nor Kripke thought that mind-brain identities were necessary; mental states are defined by their functional or phenomenal properties, and not their physical properties.

liquid' does not have a *superficial use* in which 'liquid' *flexibly* picks out all and only liquidish substances. Anyway, suppose the Semantic Internalist concedes the point - after all physical science rapidly progresses and even if 'liquid' used to have a superficial use, maybe it has now a natural kind use. It will still be true that *secondary predicates* like 'colourless', 'tasteless' and 'thirst-quenching' have superficial uses. Any substance, no matter what its micro-structure, which has the effect of quenching thirst upon being consumed by individuals is thirst-quenching. All that counts for something to be thirst-quenching is that it elicits a certain *behaviour*, e.g. ingested by individuals after physical exercise, and maybe prompts certain *phenomenological feels* - specific sensory qualities - and not how the underlying micro-physics is.¹⁴

The Semantic Externalist may try a different tack. Take this Burge-style [1979] Argument:

(A) Suppose that, in an *actual situation*, I have many true beliefs about arthritis, but also assent to 'I have arthritis in my thigh'. However, since arthritis is necessarily an ailment of the joints only, *I falsely believe that I have arthritis in my thigh*;

(B) Now suppose there is a *counterfactual situation* entirely identical to the actual except that 'arthritis' applies not only to arthritis, but also to rheumatoid ailments outside the joints, including the one in my thigh. In this situation I am also disposed to assent to 'I have arthritis in my thigh'.

(C) In (B) I cannot believe that I have arthritis in my thigh, indeed no belief-ascription containing 'arthritis' in opaque position is true of me. Instead I *truly believe that I have twin-arthritis in my thigh*, since this is what the sentence 'I have arthritis in my thigh' mean.

¹⁴ It is probable that *functional predicates* will have such superficial uses. Consider 'being a heart', 'being a clutch' or 'being a supermarket', etc. A heart is whatever fills the functional role that hearts have: pumping blood with a certain frequency, etc. In most of us hearts are carbon-based, but pacemakers are hearts if they perform the role hearts do in most of us. Or take theoretical terms like 'vitamin' or 'anaesthetics' whose functional roles are also filled by different kinds of stuff within W_A . For disagreement, see Putnam [1996, pp. 25-8] who believes that 'pencil' works just like 'water'.

Let us call (A)-(C) the *Arthritis Argument*. It has some intuitively plausible premises. (i) It is wrong to interpret (A) as a situation in which I *truly* believe that I have twin-arthritis in my thigh¹⁵. (ii) The propositional content of 'arthritis' in (A) is not identical to the propositional content of 'arthritis' in (B) since the terms are extensionally non-equivalent. (iii) Beliefs and other cognitive attitudes are individuated by the truth-conditions of their content-clauses.

According to Burge [1979, 1982], the Arthritis Argument establishes a kind of *Social Externalism*: mental content is individuated by *social facts about linguistic use* in one's speech community. In (B) I have all and only the intrinsic properties I have in (A), but, given (ii), the contents of my beliefs are different, and so, given (iii), I am in different belief-states. It follows that the mental property of having a belief with a certain content does not supervene on intrinsic properties of individuals. Where the Twin Earth Argument moved from subjectively indistinguishable differences in *the physical environment* to differences in semantic and mental content, the Arthritis Argument goes from subjectively indistinguishable differences in *the social environment* to differences in semantic and mental content. But, if good, the Arthritis Argument will apply right across the language, and the upshot would be a kind of *Rampant Semantic Externalism*. Consider this variant of (A)-(C):

(A') Suppose that, in an *actual situation*, I have many true beliefs about thirst-quenching such as that water is thirst-quenching, but also assent to 'rye bread is thirst-quenching'. However, since thirst-quenching is necessarily a property of liquids only, *I falsely believe that rye bread is thirst-quenching*.

¹⁵ I think Burge is right in denying that full understanding of a content is a necessary condition for believing that content. Had I understood 'arthritis' correctly, I would have had the same belief I have when having incomplete understanding of 'arthritis'.

(B') Now suppose there is a *counterfactual situation* entirely identical to the actual except that 'thirst-quenching' applies not only to thirst-quenching, but also to hunger-satisfying substances. In this situation I am also disposed to assent to 'rye bread is thirst-quenching'.

(C') In (B') I cannot believe that rye bread is thirst-quenching; indeed no belief-ascription containing 'thirst-quenching' in opaque position is true of me. Instead I *truly believe that rye bread is twin-thirst-quenching*, since this is what the sentence 'rye bread is thirst-quenching' means.

It is easy to construe similar arguments for the rest of the watery properties, and so it seems to follow that the supposedly narrow content of 'water' is inexpressible.¹⁶ The Semantic Internalist should, however, not be carried away by this. First of all, Semantic Internalism is not the claim that every speaker knows everything about every concept. Nor is it the claim that the knowledge a speaker has of a concept always suffices to determine the application-conditions for that concept. That is why we all agree on (i). So, why are (A)-(C) and (A')-(C') not just cases in which a speaker has an incomplete grasp of the relevant concepts? Because this point is merely a premise in the arguments: what it shows is that if understanding is incomplete, then understanding is not as the Semantic Internalist says it is. The real question is why we should concede that the content, hence the truth-value, of my belief changes when the language changes. When we ascribe cognitive attitudes, we should always strive to make our fellow speakers as rational as possible. But why should I all of a sudden be right about the world just because a linguistic convention

¹⁶ Burge himself mentions 'brisket', 'mortgage', 'red', and 'sofa'. Lewis [1999b, p. 314], however, disavows Rampant Semantic Externalism: "we should not jump to the conclusion that just any belief sentence is susceptible to Twin Earth examples. Oscar thinks that square pegs don't fit round holes; I don't think you can tell an even halfway convincing story of how Twoscar, just by being differently acquainted, fails to think so too." But what is to stop us from setting up an Arthritis Argument on mathematical terms like 'square'? In an *actual situation*, I have the *true* belief that *square pegs don't fit round holes*, but in a *counterfactual situation* in which 'square' applies to round as well as square objects, I would have the *false* belief that *twin-square pegs don't fit round holes* - even if I remained intrinsically identical. The premise about partial understanding thus seems inessential to the Argument.

has changed behind my back? Thus maybe more sense can be made of my behaviour if, say, in both (A) and (B) I am ascribed the *true belief that I have a rheumatoid ailment in my thigh*. This ascription involves no change in use of language and 'arthritis' is not used in opaque position. To my knowledge, Burge has not argued why this possibility can safely be ignored. The response would no doubt be to run the Arthritis Argument once again on 'rheumatoid ailment'. In addition, the Semantic Externalist could via appeal to Disquotation (D) insist that we should remain faithful to my - sincere and reflective - assent to 'I have arthritis in my thigh'.

Compare with Twin Earth. One could consistently maintain that what I truly believe when I assent to 'water is wet' is that the watery stuff is wet, whether I am on Earth or on Twin Earth. But in this case, it is more plausible that my belief would change, had I been on Twin Earth, since after all the world is different on Twin Earth. What should be born in mind is that, to repeat, what we learn from Twin Earth is not that reference cannot go via associated properties. On the contrary, we learn *which* properties mediate reference: water is the watery stuff of my acquaintance. The same is true of the Arthritis Argument. In general, I am unaware that a referring term F applies to objects o only, and not to objects o*. I actually have the false belief that o* is F, but, for all I know, I might have had the true belief that o* is twin-F, had F applied to both o and o*. Since the use-facts which actually and counterfactually govern F are external to me, they may change while I remain intrinsically the same. But if the use-facts change, then the content of my belief changes, and so the property of having a belief with that content fails to supervene on my intrinsic properties. Nevertheless, some expert-speakers do know the relevant use-facts: F applies to o only, 'arthritis' applies to ailments of the joints only, 'thirst-quenching' applies to liquids only, etc. We can therefore insist that the corresponding descriptions are part of the cluster associated with those terms: by 'arthritis' I associate the property that goes under the name

'arthritis' in my language community, etc.¹⁷ So, when I use 'arthritis' in (A), I refer to what my speech community refers to in the actual situation, namely ailments of the joints only. But when I use 'arthritis' in (B), I refer to what my speech community refers to in the counterfactual situation, namely ailments of the joints as well as outside the joints. But given that I remain intrinsically the same in the two contexts of acquisition (A) and (B), the predicate 'being arthritis' is wide in that its content is context-dependent. What the Arthritis Argument shows is that associated with a term *F* is the *other-dependent* description '*being referred to as F by my speech community*'. What happens when a speaker's understanding is imprecise is that she relies upon those expert speakers who do know the relevant description '*F* applies to *o* only'. It can thus happen that a speaker associates *inaccurate* descriptive information with *F*, and yet succeeds in using *F* to pick out its referent. Similarly, it may be that a speaker does not associate *enough* descriptive information with *F* to identify its referent uniquely, and yet is able to use *F* to pick out its referent. E.g. Putnam's [1996] alleged inability to distinguish elms from beeches does not prevent him from using 'elm' to refer uniquely to elm-trees. Of course Putnam knows how to disquote, and so he knows that only elms are called 'elm' in his language, but that will not put him in a position to discriminate between elms and beeches. In this case, Putnam must defer to his fellow speakers who do possess sufficient and accurate information.¹⁸ So, just as we have

¹⁷ The two cases mentioned should be distinguished. In the case of 'arthritis' the speaker has an imperfect grasp of certain technical application conditions and must thus consult the experts, e.g. her doctor. Here we have a genuine case of what Putnam [1996] called *the division of linguistic labour*. Since such conditions cannot be expected to be known by ordinary competent speakers, they do not enter into the cluster of associated descriptions. In such cases, my competence is parasitic on the experts' competence. In the case of 'thirst-quenching', however, the speaker has a partial understanding of non-technical application conditions, and so merely defer to what every non-expert competent speaker knows. It is therefore more plausible in this case that the condition is part of the cluster, i.e. is part of what someone knows who has full competence with 'thirst-quenching'.

¹⁸ It is often argued against Descriptivism that such cases demonstrate that identifying knowledge is neither necessary nor sufficient in order to be competent with a singular term. It seems to me, however, that

seen that *causal* descriptions - 'the watery stuff of our acquaintance' - involve wide content, so will *buck-passing* descriptions - 'the ailment referred to as 'arthritis' by our speech community'. Consequently, just as we could factor out the narrow property of being acquainted with a watery stuff, we can factor out the narrow property of having an ailment called 'arthritis' by a speech community.

It is obvious that a satisfactory defence of narrow content would demand more substantial argument than I have canvassed here. Also there are other objections that I have not dealt with.¹⁹ But at least the foregoing is suggestive of how the defence might go. In the last three Sections of this Chapter I shall return to the problem about Incompatibilism as presented in Chapter 1. But first we must go back to Dry Earth.

5.4. The Empty Case Revisited

According to Referentialism, an utterance of a simple sentence is assigned singular truth-conditions which specify the conditions under which the sentence is true-in W_A and true-of all W_P . According to Descriptivism, however, an utterance of a simple sentence is assigned rigidified descriptive truth-conditions which specify the conditions under which the sentence is true-in all W_P and true-of all W_P . I have argued that rigidified descriptive truth-conditions are a priori knowable since they do not depend on which W_P is in fact actual and

understanding comes in levels, and that there is an expert level at which Putnam is incompetent with 'elm', hence that he lacks *full* understanding of 'elm'. The reason Putnam can lack unique identifying knowledge and still pick out all and only elms by 'elm' is that he relies on the experts who do attach properties to 'elm' that *inter alia* discriminate between elms and beeches. See also Stanley [1999b].

¹⁹ Another worry put forward by Stalnaker [1999a, pp. 177-8, 199-200] and Block [91, p. 40] is that *narrow content is too narrow*. We want narrow content to be shared by real people and not just intrinsic duplicates. But it is difficult to see how our definition should be relaxed, and in particular, how talk of near duplicates would resolve in the appropriate thickening of the notion of narrow content. I do not have a worked-out proposal as to how it might go. Suffice it to say that spelling out narrow content in terms of a supervenience-relation allows for the possibility that two individuals who share narrow content differ in intrinsic properties.

so can be known independently of any particular empirical facts about W_A . Descriptive content is independent of any particular context of utterance or acquisition. Singular content, however, is a posteriori knowable since it is dependent on which W_P is in fact actual, and so can only be known once the relevant empirical facts are in. Moreover, in Sec. 5.1 I introduced a threefold distinction between object-dependence, context-dependence and world-dependence. In Sec. 5.3 I argued that no content is world-independent: I believe that the watery stuff is wet, but my doppelgänger in a far-fetched W_P may fail to believe just that. Yet I argued that such purely descriptive content is context-independent: I believe that the watery stuff is wet and so does my doppelgänger on a Twin Earth in W_A . Purely descriptive content is shared by doppelgängers within the same W_P . I have also introduced the idea that some content is context-dependent yet object-independent. I shall try to develop this idea in a short while. But for now I wish to focus on the object-dependence of singular content. We saw in Sec. 1.6 that all the Twin Earth Argument shows, if good, is that content is not individuated individualistically: if doppelgängers are embedded in different physical environments, then the contents of their beliefs differ. That is, not all content-bearing mental properties are shared by intra-world doppelgängers. But the Twin Earth Argument does not speak to the empty case. Could my doppelgänger share my beliefs were he embedded in an environment where the relevant external facts go missing? In other words, it is one thing to say that which mental contents I entertain depend on which external objects I interact with; it is another thing to say that the very existence of those contents depend on the existence of those objects. In Sec. 1.6 I called the former *Weak Semantic Externalism* and the latter *Strong Semantic Externalism*, and I followed Boghossian [1997, 1998] in arguing that Weak Semantic Externalism was an unstable position liable to collapse into its stronger counterpart. The question we should now address is whether one can hold merely Weak Semantic Externalism with respect to singular content. I think it is easy to see that this question should be answered negatively. Consider again Dry Earth which is like Earth in every respect except that we are all deluded

into thinking that there is a stuff which has all and only the watery properties.²⁰ Despite all appearances, nothing actually fills the oceans, lakes, rivers, etc. Incredibly enough, all tokens of 'water' are simply empty. The question is: what are the singular truth-conditions of Dry Earthly tokens of sentences containing 'water'? On Earth utterances of sentences containing 'water' have truth-conditions which are singular with respect to H_2O , and on Twin Earth they have truth-conditions which are singular with respect to XYZ. But with respect to which kind of stuff do utterances on Dry Earth of sentences containing 'water' have singular truth-conditions? An utterance of 'water is wet' is true iff ? is wet. It is clear that there is no answer since there is no watery stuff on Dry Earth with respect to which they could be singular. As used on Dry Earth 'water' does not pick out any kind of stuff that could feature on the right-hand side of the bi-conditional. So, not only are all 'water'-sentences false, or lack truth-values, on Dry Earth, there are even no conditions under which they would have been true. So, it seems that singular content is not only not individuated individualistically, i.e. context-dependent, it is also object-dependent.

This result poses an immediate threat to Referentialism according to which the sole propositional content of a sentence containing 'water' is a singular proposition. The Referentialist cannot simultaneously subscribe to these three tenets: (i) Strong Semantic Externalism; (ii) We would understand utterances of 'water'-sentences were we to find ourselves on Dry Earth; (iii) Linguistic understanding is of semantic, i.e. truth-conditionally relevant, content. Here is why. Were I to find myself on Dry Earth, I would still understand 'water'-sentences, hence have knowledge of their propositional content. According to Strong Semantic Externalism that means I have knowledge of the singular propositions expressed. But *ex hypothesi* no such singular propositions are expressed on Dry Earth. So, either (i), (ii) or (iii) must go. Which one? Well, we have just argued that (i) must hold given Weak Semantic Externalism which the Referentialist no doubt is happy to endorse.

²⁰ It does not matter for present purposes whether we take Dry Earth as a remote planet in W_A or as a W_P qua actual. What is important is that Dry Earth is considered as a possible context of acquisition.

So, maybe we do not really understand 'water'-sentences on Dry Earth because there is nothing to understand. On Dry Earth no singular truth-conditions can be pinned down, so if understanding is of truth-conditional content, then no understanding is forthcoming! That, however, seems incredible. Consider this example. Suppose my doppelgänger on Dry Earth is thirsty and utters 'that glass contains water'. He reaches out for the glass but in vain. Clearly his utterance has significance in that it explains his behaviour. Given his desire to quench his thirst, his behaviour makes sense if we can understand him to be expressing a belief with 'that glass contains water'. But if Dry Earthians suffer from a grand illusion of content, then my doppelgänger fails to express a belief with singular content by his utterance, and so there can be no belief which in conjunction with his desire is causally responsible for his behaviour.²¹ Or just consider mundane cases of reference failure. Not knowing that I suffer from an illusion of content, I point to a bush while saying 'that man is drunk'. Since there is no one there, my utterance has no singular truth-conditions. But clearly my companion understood what I said, namely that I had the false, but causally efficacious, belief that there was a drunk man in the bush - I went to have a look in the bush.

So, we are left with (iii). Is it possible that understanding is not truth-conditional? Putnam advocated this view in e.g. [1978] where he subscribed to a two-component theory of natural kind concepts. On the one hand there is the contribution from the environment: the reference of 'water' is fixed by certain in principle defeasible criteria used by experts.

²¹ Illusions of content provide additional support for the existence of narrow content. Suppose my doppelgänger and I are both thirsty and that we both have a belief we would express with 'that glass contains water'. The only difference between us is that my doppelgänger suffers from an illusion of content. We perform the same bodily movement of reaching out for what we think is a glass of what we call 'water', but only I succeed. My doppelgänger is not irrational, so presumably something explains his behaviour. But what? Not the wide content belief that the glass contains water since he has no such belief. So perhaps the narrow content belief that the glass contains a watery stuff explains his behaviour. But if the narrow content belief suffices to explain his behaviour, then it also suffices to explain my behaviour - on the plausible assumption that no content is available to him that is not available to me.

On the other hand there is the stereotype: standardised set of idealised beliefs commonly associated with 'water' to facilitate communication. Putnam was keen to separate out a theory of meaning for 'water' which pertains to the classificatory extension of 'water', and a theory of understanding for 'water' which concerns communal beliefs and interaction. What is required to be competent in the use of 'water' is grasp of certain descriptive stereotypes, but what determines the meaning of 'water' are facts about which kind of physical stuff 'water' picks out. He thus eschewed the Dummettian dictum that *meaning and understanding are correlative notions*. The semantics is all provided by the theory of meaning since the stereotypical beliefs are unfit to get the reference of 'water' right in various actual and counterfactual circumstances.

Now, I have argued that we should endorse a theory of singular terms which only acknowledges descriptive content. Purely descriptive content determines truth-conditions for all W_P qua actual, and rigidified descriptive content determines truth-conditions for all W_P qua counterfactual. In light of those arguments, I think we are better off upholding Dummett's dictum: knowledge of purely descriptive content yields a grade-one understanding and knowledge of rigidified descriptive content yields a grade-two understanding. Let us therefore see how Descriptivism handles empty cases. We have seen that singular thoughts are *relational* in the sense of owing their existence to the existence of the objects they essentially are about. Descriptive thoughts, however, are not relational since they are not essentially about any particular objects. On Earth, my utterance of 'water is wet' is true on Earth iff the watery stuff is wet. But the same utterance would have been true under the same conditions had it taken place on Twin Earth or on Dry Earth. On Twin Earth, the sentence would have been true due the fact that XYZ is wet, and on Dry Earth, the sentence would have been false due to the fact that, as it turns out, nothing has enough of the watery properties to deserve the name 'water'. One can lay down descriptive truth-conditions independently of whether one is embedded on Earth, Twin Earth or Dry Earth. This reflects the fact that descriptive thoughts are *satisfactional* in the sense of being about

if anything, then whatever satisfies enough of certain associated conditions. They do therefore not have the particular object-dependency that singular thoughts have. So, according to Descriptivism, understanding is truth-conditional even on Dry Earth. Moreover, given that we are not conceptually confused in the way the Referentialist thinks we are, it is easy to account for the causal power of mental properties on Dry Earth. The reason my doppelgänger reaches out for a glass to quench his thirst is that he (falsely) believes it contains a watery stuff. Or take the man in the bush. Since there is nobody there, no singular truth-conditions can be pinned down for my utterance. In this case, we can fall back on descriptive truth-conditions: 'that man is drunk' is true iff the demonstratively identified male is drunk. The reason I inspect the bush is thus that I (falsely) believe there is a male person in it who is drunk (and I desire to help drunks or whatever).²²

5.5. A Response to the MC-form

We are now in a position to diagnose and respond to the Incompatibilist Argument from Sec. 1.5. *Incompatibilism*, remember, was the claim that *Semantic Externalism* and *Privileged Access* are incompatible doctrines. It is easy to see a *prima facie* tension: how can we have introspective access to the contents of our occurrent mental states if those contents depend for their individuation, indeed for their existence, on external facts that we have no introspective access to? I argued in Sec. 1.4 that the Semantic Externalist has offered no satisfactory account of why we should not think that this *prima facie* tension goes much deeper. In particular, Burge's model of self-verifying thoughts would at most

²² This is a general point about reference failure. I used to believe that my dad was stronger than Superman. In quarrels with my mates I would say: 'my dad can beat Superman!'. But if Referentialism is true, why did my mates care to disagree? There should be nothing I believed except falsely that 'my dad can beat Superman' expressed a proposition. Things look different, however, if what I falsely believe is that there is a masked super-hero who flies around and chases criminals and who is such that my dad can beat him.

show that in a very limited number of cases, Semantic Externalism and Privileged Access are consistent doctrines. But fully-fledged compatibility is another matter and, to my knowledge, no Semantic Externalist has as yet ventured to address this task. We can now see why. Wide content in its singular guise is simply not subject to Privileged Access. Bear in mind that Privileged Access is the thesis that a competent speaker can have a priori knowledge of the contents of his occurrent mental states, where by 'a priori' we mean that those contents are reflectively accessible to S - independently of any empirical investigation of his external environment. But we have seen that speakers do not have a priori knowledge of the singular contents of their wide content mental states. Singular content is knowable only after the relevant empirical information about the context of utterance or acquisition is in. I have spent some time arguing for this in the case of natural kind terms but the point is general. Suppose I leave a note on my office door saying 'back in five minutes'. Mary does not know 'its' singular truth-conditions unless she knows when I wrote it. What she knows a priori is a set of conditionals of the form: if I wrote the note at 3, then 'back in five minutes' is true iff I am back at 3.05, etc. Or suppose Mary is in the bush. I do not know the singular truth-conditions of the utterance 'she is in the bush' unless I know that Mary is in the bush. What I know a priori is that if Mary is the person in the bush, then 'she is in the bush' is true iff Mary is in the bush, etc. Such conditionals can be pinned down if one has knowledge of descriptive or linguistic content since it determines how actual and possible contexts determine which singular contents are expressed. So, what one knows a priori is the descriptive content associated with expressions since it remains constant across actual and possible contexts of utterance or acquisition.

Let us now return to the Incompatibilist Argument from Sec. 1.5. That Argument was a *reductio* of the conjunction of Strong Semantic Externalism - the claim that mental content is object-dependent - and Privileged Access:

- (1) I have mental property M,
- (2) If I have mental property M, then I meet condition C,

(3) So, I meet condition C.

This MC-form is deductively valid. Moreover, the first premise is a priori knowable by Privileged Access; the second premise is a conceptual truth, hence a priori knowable, according to Strong Semantic Externalism; hence the conclusion should be a priori knowable. But it is manifestly not a priori knowable. Take M to be the mental property of having the belief that water is wet and take the external condition C to be that water exists. Plug these ingredients into our recipe and we get:

(1*) I believe that water is wet,

(2*) If I believe that water is wet, then water exists,

(3*) So, water exists.

Privileged Access tells me that I can know a priori that I occurrently believe that water is wet. Strong Semantic Externalism tells me from the armchair, as it were, that if I believe that water is wet, then water exists. To set up thought experiments concerning the possession conditions for such beliefs on Earth, Twin Earth or Dry Earth is not to carry out an empirical inquiry. But I do clearly not know a priori that water exists since I do not know a priori that I am not on Twin Earth or on Dry Earth. Bear in mind what actually happened in the case of caloric or phlogiston.

Should this Argument carry conviction against the Referentialist who, remember, holds that the content of my belief that water is wet is singular? Well, we have seen in Sec. 5.4 that singular content is object-dependent, and so (2*) is indeed a priori knowable. We know from a thought-experiment that had I been on Dry Earth, I could not have had the singular content belief that water is wet. Earthly tokens of 'water is wet' express the singular proposition that H_2O is wet, and so not a proposition Dry Earthlings can express: 'water is wet' is true iff ?-stuff is wet. But (1*) is clearly not a priori knowable. Singular content is at most a posteriori knowable: I cannot know a priori that I believe that water is wet if the content of that belief is singular. Grasp of singular content requires knowledge of the micro-physics of W_A , i.e. knowledge of which underlying chemical kind actually

instantiates the watery properties. One can know a priori that if H_2O is the watery stuff, then 'water is wet' is true iff H_2O is wet, but there is no a priori way of affirming the antecedent. So, according to Referentialism, there is no a priori warrant for the first premise that can be transmitted across the a priori knowable conditional in the second premise to the conclusion.

How does it look on the Descriptivist's account? Well, we have argued that knowledge of descriptive content is a priori knowable. I know a priori that if water exists, then water is the watery stuff. So, if by 'water is wet' in (1*) we mean the descriptive proposition associated, then indeed (1*) is a priori knowable. But if thus taken, (2*) is simply false. Descriptive propositions are not object-dependent. Dry Earthly tokens of 'water is wet' are true iff the watery stuff is wet. (2*) is only true if by 'water' we understand the singular content expressed. So, according to Descriptivism, there is indeed an a priori warrant for the first premise; it is just that there is no a priori warranted conditional across which it could be transmitted. To be sure, the conditional in the second premise is not even a posteriori knowable; it is false.

So, here is a *dilemma* for the Incompatibilist. If one starts off making the first premise a priori knowable, then the second premise comes out false, but if one starts off making the second premise true, then the first premise comes out at most a posteriori knowable. So, it would seem that if ambiguity is to be avoided, then there is no way both premises can simultaneously be rendered true a priori. So, the MC-form is blocked: there is no non-equivocal way of reaching the conclusion that we have a priori knowledge of the external condition C.

The Incompatibilist may object to the foregoing response that it simply misses the point. The MC-form was set up as a *reductio* of Compatibilism which is the claim that we can attain Privileged Access to the kind of wide content that Strong Semantic Externalism subscribes to. The Referentialist account has not vindicated Compatibilism. On the contrary, it seems to concede the Incompatibilist contention that no Privileged Access to

wide content mental states is forthcoming. What we have shown is that the first premise is true of descriptive content, whereas the second premise is true of singular content, but the Incompatibilist need not demur to this. Moreover, it seems the MC-form still stands. Suppose we set out to render the second premise true, i.e. take the mental property M to be singular content mental properties. Then it is still true that were we to have Privileged Access to our singular content mental properties, then we would be able to have a priori knowledge of the external condition E. This *counterfactual* claim has not been contested but, the thought goes, it is what the Compatibilist claims to be true in *actual* fact. So, according to the Compatibilist, nothing should stop us from achieving armchair knowledge of ordinary, empirical state of affairs that we would otherwise only have knowledge of were we to engage in some successful empirical investigation.

This of course is all to the point. In light of the foregoing, it is more than hard to see how Strong Semantic Externalism could possibly be reconciled with Privileged Access. Presented with the MC-form, the friend of Strong Semantic Externalism has no choice but to deny that we have Privileged Access to object-dependent content. So, the Incompatibilist is clearly right that one cannot at the same time subscribe to both tenets. And the significance of this fact is that if one holds, as the Referentialist does, that mental content is singular, then one better explain why it is perfectly harmless to surrender the a priority of our knowledge of our own occurrent mental states. In any case, the Descriptivist is not doomed to struggle with Incompatibilist arguments. On his account, the second premise in the MC-form is simply false. What we shall now finally see is that Descriptivism paves the way for a kind of Neo-Compatibilism according to which some wide content is subject to Privileged Access without the threat of paradox.

5.6. Beyond Incompatibility

Semantic Externalists and Semantic Internalists alike have accepted the conditional claim that *if Incompatibilism is true, then Semantic Externalism is false*. The Semantic Internalists have therefore invested much energy arguing why the antecedent is true and the Semantic Externalists have spent equally much energy trying to rebut the Semantic Internalists' arguments.²³ The reason for this is obviously that Privileged Access is a forceful principle: in the normal run of things I know the contents of my occurrent thoughts just by introspective reflection. Now, if Semantic Externalism goes hand-in-hand with Referentialism, then indeed I suspect much energy has been wasted trying to prove Incompatibilism wrong. But we have learned from the Twin Earth Argument and the Arthritis Argument that there is space for a weaker notion of wide content than the kind of object-dependent content which is entailed by Referentialism.

Consider first the Twin Earth Argument. We started off with the assumption that the content of 'water' was purely descriptive: water is the watery stuff. The watery stuff on Twin Earth, however, is not stuff we would describe as 'water' since it is not identical to the watery stuff *around here*, as Putnam put it. So, what we learned was that the cluster of watery properties associated with 'water' must include *causal properties*: water is the watery stuff of our acquaintance. The consequence was that certain mental properties like water-beliefs fail to supervene on intrinsic properties. My utterance of 'water fills the oceans' is true iff the watery stuff of *our acquaintance* fills the oceans. But this is not the condition under which my doppelgänger's utterance of the same sentence is true on a remote Twin Earth in W_A . His utterance is true iff the watery stuff of *their acquaintance* fills the oceans.²⁴ My doppelgänger and his fellow speakers have had no causal encounters

²³ See for instance Burge [1996], Davidson [1996], Brueckner [1998] and McLaughlin & Tye [1998].

²⁴ The reason I use 'our acquaintance' instead of 'my acquaintance' is that it need not be me who is acquainted with water but other members in my speech community. The important point is not to read 'our' as referring to both me and my doppelgänger. He is not acquainted with the watery stuff of our acquaintance - H_2O - since

with the watery stuff I and my fellow speakers interact with. So, my belief that water fills the oceans is context-dependent in that my doppelgänger in a different context of acquisition - Twin Earth - fails to share my belief. But the content of my belief is not object-dependent. On Dry Earth, my utterance of 'water fills the oceans' still express the causally constrained descriptive proposition that the watery stuff of our acquaintance fills the oceans. Since there is no watery stuff that I interact with on Dry Earth, there is nothing that makes that belief true, but there are still determinate conditions under which it would have been true. Adding a causal property to the cluster term 'water' does not make its associated content object-dependent since it is just another way of constraining its reference without saying which kind of stuff it picks out. Although my belief that water fills the oceans is thus individuated by reference to facts in my physical environment, it is nevertheless a belief whose content I have Privileged Access to. I can know a priori that if water exists, then water is the watery stuff. All that takes is reflection on how to describe various W_p were they actual. But if I can know that, I can also know a priori that if water exists, then water is the watery stuff of our acquaintance. All that takes is reflection of how to describe Twin Earth. The same is true of my doppelgänger. He will also associate a causal property with 'water' which he will express by 'the watery stuff of our acquaintance'. But given that he inhabits a different physical environment, 'water' will in his mouth pick out a different kind of stuff. So, we can both have the property of being acquainted with a watery stuff, but only I can have the property of being acquainted with the watery stuff of our acquaintance. So, the upshot is that causally constrained descriptive content is wide in the sense of being merely context-dependent yet is subject to Privileged Access.

The same is true of the Arthritis Argument. What we learned from that Argument was that we must include *other-dependent* properties as part of the cluster associated with a

he belongs to a speech community whose members are acquainted with a different watery stuff - XYZ. Alternatively, the same point can be brought out by deploying a distinction between the indexicals *around here/around there*.

singular term. What I refer to by, say, 'water', and hence what I mean, is not only a function of the properties I associate with 'water', it is also dependent on which properties my speech community attach to 'water'. In my use of 'water', I must defer to my fellow speakers' use of 'water': water is the watery stuff that is *called 'water' by my speech community*. This means that if my doppelgänger inhabits a different speech community, then he will associate a different property with 'water', namely the property of being *called 'water' by his speech community*. His use of 'water' does not answer to the way 'water' is used in my community. He will of course use the same sentence 'water is the stuff that goes under the name 'water' in my speech community', but given that he is in a different context of acquisition, 'water' will pick out a different kind of stuff.²⁵ So, the truth-conditions of our respective utterances of sentences containing 'water' will be different, and so my belief that water fills the oceans fails to supervene on my intrinsic properties. Although the content of this belief is thus individuated by reference to facts in my social environment, it is a belief whose content is subject to Privileged Access. All it takes to realise that we must associate other-dependent properties with singular terms is reflection on how to describe circumstances in which my doppelgänger and I are embedded in different speech communities. So, the upshot is that other-dependent descriptive content is wide in the sense of being context-dependent yet is subject to Privileged Access.

There is thus a striking analogy between the Twin Earth Argument and the Arthritis Argument in that both emphasise the need for *indexical properties* as part of the cluster associated with singular terms. Consequently, the contents of sentences containing such terms will be sensitive to the context in which they are uttered. The content of my belief that water is fun to swim in is dependent on the context in which I acquired that belief due to the fact that with 'water' is assigned the property of being the stuff *around here* that *my speech community* calls 'water'.

²⁵ Again we could factor out a property my doppelgänger and I both associate with 'water', namely the property of being referred to as 'water' by *a* speech community.

It may be objected that indexical properties entail differences in content, not only when we want them, but also when we do not want them. Suppose *Perfect Twin Earth* is, not a near duplicate of Earth like Twin Earth, but an exact duplicate of Earth. Perfect Twin Earth is a place located on the other side of the galaxy where the watery stuff is H_2O . Now, just as on Twin Earth, when my Perfect Twin Earth doppelgänger and I both utter the sentence 'water is fun to swim in', what we say have different truth-conditions due to the context-sensitivity of 'here', 'my', etc. But the intuition is that since we are intrinsically identical, speak the same language and inhabit *identical* physical environments, he believes just what I believe when I announce 'water is fun to swim in'. Moreover, if my use of, say, 'around here' excludes Twin Earth, then it would be *ad hoc* to think it includes Perfect Twin Earth; suppose the latter is more remote than the former. Therefore, the thought goes, our concept of water cannot involve covert reference to my environment.²⁶

I think there are two distinct questions. One is whether the stuff on Perfect Twin Earth is water; another is whether my doppelgänger believes what I believe. Given that the stuff called 'water' on Perfect Twin Earth is both watery and H_2O , it is intuitively also stuff we would call 'water'. But how can that be if our tokens of 'water' pick out the watery stuff *around here*? Well, bear in mind that, according to Descriptivism, 'water' refers to whatever stuff satisfies *sufficiently many* of the properties that make up the cluster associated with 'water'. We have seen that 'water' may refer to actual samples of H_2O even if they lack one or more of the purely descriptive properties. I think Perfect Twin Earth highlights the possibility that we may use 'water' to refer to actual samples of H_2O even if they lack certain indexical properties such as being the prevalent liquid around here. We should allow for the discovery of samples of water in W_A which we have had no causal encounters with. The cluster is not a list of necessary and sufficient conditions, and there is no reason to regard any single property as sacrosanct. What is more, we are willing to call those samples 'water' only in virtue of the fact that they are of the same kind - H_2O - as the

²⁶ This objection can be found in Pryor [2000, pp. 108-22].

stuff to which we do bear certain causal relations and which does have those watery properties. We would not have labelled the watery stuff on Perfect Twin Earth 'water' had it not been of the same kind as the watery stuff of our acquaintance on Earth. So, my doppelgänger clearly believes *of* water that it is fun to swim in. But I suspect that even if there is a sense in which we share De Dicto beliefs, e.g. the belief that the watery stuff is fun to swim in, the difference in location will force a distinction in what we believe. Given that he has no beliefs about Earth, the belief he expresses by 'water is fun to swim in' should concern only Perfect Twin Earth samples of water; after all, those are the samples he, or his community, has interacted with. Similarly, if his grasp of 'water' is somewhat distorted, he does not defer to speech communities located on far-off planets even if they share his language; he answers to the experts in his own community.²⁷

Let me sum up. The good news is that once Semantic Externalism is divorced from Referentialism, there is hope for a kind of Neo-Compatibilism. We learn from the various arguments against Descriptivism and Semantic Internalism which properties must be associated with singular terms, and although some of them will issue in wide content mental states, they will all be reflectively accessible properties. Nevertheless, we can still hope to factor out narrow components. I have tried to argue that the prospects for a thorough-going Semantic Externalism look rather gloomy. It is not to be expected that the argumentative strategy which, say, the Twin Earth Argument deploys - change the world while fixing the way it appears - is going to give us wide content across the board. For a wide range of *phenomenal* and *functional concepts*, there will be no relevant distinction

²⁷ The same is *mutatis mutandis* true of 'actually'. Thus Fitch [1981] and Soames [1998] have argued that, say, 'water' cannot be short for the *actual* watery stuff because I have doppelgängers in counterfactual W_P who have beliefs about water yet no beliefs about W_A . I think they are wrong. The reason why the watery stuff in such W_P is truly described by us as 'water' is that it is of the same kind - H_2O - as the watery stuff *in* W_A . It is only in virtue of indexicals like 'around here' or 'actual' that we describe speakers who have no beliefs about our vicinity as having beliefs about water or whatever.

between the way the world is and the way the world appears, that is, they will not be referentially sensitive to variations in the underlying physics of the world.

Bear in mind Putnam's dictum that *meaning just ain't in the head*. What I am proposing is that it be taken at face value when it comes to singular content: once meaning has been expelled from the mind, as it were, there is no prospect of attributing to meaning all the intentional properties it would have had, had it been firmly entrenched in the mind. When it comes to descriptive content, however, the dictum is false. I have argued that no convincing case has been made that singular terms cannot have associated descriptive content. The conjunction of the Modal Argument, the Semantical Argument and the Epistemological Argument falls short of a knock-down refutation of Descriptivism. On the contrary, considerations about the behaviour of singular terms in intentional contexts strongly suggest that they better have semantically relevant descriptive contents. Moreover, neither the Twin Earth Argument nor the Arthritis Argument shows that no mental content can be narrow in the sense of being context-independent though we should not expect it to be world-independent. What we learn is that some mental content will be wide in the sense of being merely context-dependent. No compelling case has been made that we must also take on board the stronger notion of object-dependent content. Both context-independent and context-dependent content are, however, firmly in the mind if that means subject to Privileged Access. This, I propose, is what it is to have meaning in mind.

Bibliography

Almog, J., Perry, J. & Wettstein, H. (eds.): *Themes from Kaplan*, Oxford, Oxford University Press, 1989.

Blackburn, S., (ed.): *Meaning, Reference and Necessity*, Cambridge University Press, Cambridge 1975,

Boghossian, P.: 'Content and Self-Knowledge', *Philosophical Topics*, XVII/1, 1989, pp. 5-26.

—————: 'The Transparency of Mental Content', *Philosophical Perspectives*, 8, Logic and Language, 1994.

—————: 'What the Externalist can Know Apriori', *Proceedings of the Aristotelian Society*, 1997, pp. 161-75.

—————: 'Replies to Commentators', *Philosophical Issues*, 9, 1998, pp. 253-6.

—————: & Peacocke, C. (eds.): *New Essays on the A Priori*, Oxford University Press, Oxford, 2000.

Block, N.: 'What Narrow Content is Not' in *Meaning in Mind*, (eds.) B. Loewer & G. Rey, 1991, pp. 33-64.

Block, N. & Stalnaker, R.: 'Conceptual Analysis, Dualism, and the Explanatory Gap', *The Philosophical Review*, 108/1, 1999, pp. 1-46.

Braddon-Mitchell, D. & Jackson, F.: *Philosophy of Mind and Cognition*, Blackwell, Oxford, 1996.

Brueckner, A.: 'What An Anti-Individualist Knows A Priori' in *Externalism and Self-Knowledge*, (eds.) Ludlow, P. & Martin, 1998, pp. 197-205.

—————: 'The Difficulties in Generating Scepticism about Knowledge and Content', *Analysis*, 59 (1), 1999, pp. 59-63.

—————: 'Ambiguity and Knowledge of Content', *Analysis*, 60 (3), 2000, pp. 257-260.

Burge, T.: 'Individualism and the Mental', *Midwest Studies in Philosophy*, Vol. IV, 1979, pp. 73-122.

—————: 'Other Bodies' in Woodfield, A., (ed.): *Thought and Object*, 1982, pp. 97-120.

———: 'Individualism and Self-Knowledge', in *The Twin Earth Chronicles*, (eds.) Pessin, A. & Goldberg, S., M. E. Sharpe, 1996, pp. 342-54. Originally published in *The Journal of Philosophy*, 85, 1988, pp. 649-63.

———: 'Our Entitlement to Self-Knowledge', in *Externalism and Self-Knowledge*, (eds.) Ludlow, P. & Martin, N, 1998, pp. 239-63.

Byrne, A. & Thau, M.: 'In Defense of the Hybrid View', *Mind*, 105, 1996, pp. 139-49.

Byrne, A.: 'Cosmic Hermeneutics', forthcoming, 2000a.

———: 'Chalmer's Two-Dimensional Argument against Physicalism', forthcoming, 2000b.

Carruthers, P.: 'Understanding Names', *Philosophical Quarterly* 33, 1983, pp. 19-36.

Chalmers, D.: *The Conscious Mind*, New York, Oxford University Press, 1996.

Crimmins, M. & Perry, J.: 'The Prince and the Phone Booth: Reporting Beliefs', *Journal of Philosophy*, 86, 1989, pp. 685-711.

Crossley, J. N., & Humberstone, I. L.: 'The Logic of "Actually"', *Reports on Mathematical Logic* 8, 1977, pp. 11-29.

Davidson, D. & Harman, G. (eds): *Semantics of Natural Language*, Dordrecht, Reidel, 1972.

Davidson, D.: 'Knowing One's Own Mind', in *The Twin Earth Chronicles*, (eds.) Pessin, A. & Goldberg, S., 1996, pp. 323-41. Originally published in *The Proceedings of the APA*, 60, 1987, pp. 441-58.

Davies, M. K., & Humberstone, I. L.: 'Two Notions of Necessity', *Philosophical Studies* 38, 1981a, pp. 1-30.

Davies, M.: *Meaning, Quantification, Necessity*, Routledge and Kegan Paul, 1981b.

———: 'Externalism, Architectualism and Epistemic Warrant' in *Knowing our Own Minds*, (eds.) C. MacDonald, B. C. Smith, C. Wright, 1998, pp. 321-61.

———: 'Externalism and Armchair Knowledge' in *New Essays on the A Priori*, (eds.) Boghossian & Peacocke, 2000, pp. 384-414.

Donnellan, K.: 'Reference and Definite Descriptions', *Philosophical Review*, 75, 1966, pp. 281-304.

—————: 'Proper Names and Identifying Descriptions' in Davidson, D. & Harman, G., (eds.), *Semantics of Natural Language*, 1972, pp. 356-79.

—————: 'The Contingent Apriori and Rigid Designators' in P. French, T. Uehling, and H. Wettstein, (eds.): *Contemporary Perspectives in the Philosophy of Language*, 1979, pp. 45-60.

Dummett, M.: *Frege's Philosophy of Language*, Duckworth, London, 1973.

—————: *Truth and Other Enigmas*, Duckworth, London, 1978.

—————: *The Interpretation of Frege's Philosophy*, Duckworth, London, 1981.

—————: *The Logical Basis of Metaphysics*, Duckworth, London, 1991.

Evans, G.: *The Varieties of Reference*, Oxford University Press, Oxford, 1982.

Evans, G. & McDowell, J. (eds.): *Subject, Thought and Context*, Oxford, Clarendon Press, 1986.

Evans, G.: 'The Causal Theory of Names' in his *Collected Papers*, 1996a, pp. 1-24.

—————: 'Reference and Contingency' in his *Collected Papers*, 1996b, pp. 178-213.

—————: *Collected Papers*, Clarendon Press, Oxford, 1996.

Fitch, G. W.: 'Names and the De Re - De Dicto Distinction', *Philosophical Studies*, 1981.

Forbes, G.: *Languages of Possibility*, Basil Blackwell, Oxford, 1989.

—————: 'The Indispensibility of Sinn', *Philosophical Review*, no. 4, 1990, pp. 535-563.

Francis, B.: 'Defending Millian Theories', *Mind*, Vol. 107, 1998, pp. 703-727.

Frege, G.: 'The Thought' in *Basic Topics in the Philosophy of Language*, (ed.) Harnish, R., 1994a, pp. 517-35.

—————: 'On Sense and Reference' in *Basic Topics in the Philosophy of Language*, (ed.) Harnish, R., 1994b, pp. 142-60.

French, P., Uehling, T. & Wettstein, H. (eds.): *Contemporary Perspectives in the Philosophy of Language*, Minneapolis, University of Minnesota Press, 1979,

Goldberg, S.: 'Self-Knowledge, Self-Ascription, and the Memory Argument', *Analysis*, 59 (3), 1999, pp. 212-17.

———: 'World-Ambiguity, World-Switching, and the Semantic Intensions', *Analysis*, 60 (3), 2000, pp. 260-64.

Guttenplan, S. (ed.): *A Companion to the Philosophy of Mind*, Blackwell, Oxford, 1994,

Hale, B. & Wright, C. (eds.): *A Companion to the Philosophy of Language*, Basil Blackwell, Oxford, 1999.

Harnish, R. (ed.): *Basic Topics in the Philosophy of Language*, Harvester Wheatsheaf, 1994.

Heck, R.: 'The Sense of Communication', *Mind*, 104, 1994, pp. 79-106.

———: 'Communication and knowledge: Rejoinder to Byrne and Thau', *Mind*, 105, 1996, pp. 151-6.

———(ed.): *Logic, Language, and Reality: Essays in Honour of Michael Dummett*, Oxford, Oxford University Press, 1997.

Heil, J. & Mele, A. (eds.): *Mental Causation*, Clarendon Press, Oxford, 1993.

Jackson, F. & Pettit, P.: 'Some Content is Narrow' in *Mental Causation*, (eds.) Heil & Mele, 1993, pp. 259-82.

Jackson, F.: *From Metaphysics to Ethics. A Defense of Conceptual Analysis*, Oxford University Press, 1998a.

———: 'References and Descriptions Revisited', *Philosophical Perspectives*, 12, *Language, mind, and Ontology*, 1998b, pp. 201-18.

———: 'Representation, Scepticism, and the A Priori' in Boghossian & Peacocke (eds.), *New Essays on the A Priori*, 2000, pp. 320-32.

Kallestrup, J.: 'Puzzles about the Contingent A Priori', *Nordic Journal of Philosophy*, Vol. 1., 2000, pp. 37-47.

Kanger, S. & Öhman, S. (eds.): *Philosophy and Grammar*, Reidel Publishing Company, 1981, pp. 79-100.

Kaplan, D.: 'Demonstratives' in Almog, J., Perry, J., and Wettstein, H. (eds.), *Themes from Kaplan*, 1989, pp. 481-563.

———: 'On the Logic of Demonstratives', *Journal of Philosophical Logic*, 8, 1978, pp. 81-98.

Kim, J.: 'Supervenience' in Guttenplan, S. (ed.): *A Companion to the Philosophy of Mind*, 1994, pp. 575-83.

Kripke, S.: 'A Puzzle about Belief' in *Basic Topics in the Philosophy of Language*, (ed.) Harnish, R. 1994, pp. 352-92. Originally published in (ed.) Margalit, *Meaning and Use*, Dordrecht, Reidel, 1979.

———: *Naming and Necessity*, revised edn., Oxford, Basil Blackwell, 1980.

Lewis, D.: 'How to Define Theoretical Terms' in *Journal of Philosophy*, 67, 1970, pp. 427-46.

———: 'Possible Worlds', in his *Counterfactuals*, Harvard University Press, 1973.

———: 'De Se Attitudes and De Dicto Attitudes' in *Philosophical Review*, LXXXVIII, 1979, 513-43.

———: 'Index, Context, and Content' in *Philosophy and Grammar*, Kanger, S. and Ohman, S., (eds.) 1981, pp. 79-100.

———: 'Mad Pain and Martian Pain', in his *Philosophical Papers*, Vol. 1, Oxford University Press, 1983, pp. 122-31.

———: 'Psychophysical and Theoretical Identifications' in *The Nature of Mind*, (ed.) Rosenthal, D., 1991, pp. 204-10.

———: 'Putnam's Paradox', in his *Papers in Metaphysics and Epistemology*, Cambridge University Press, 1999a, pp. 56-78.

———: 'Reduction of Mind' in his *Papers in Metaphysics and Epistemology*, Cambridge University Press, 1999b, pp. 291-324.

Loewer, B. & Rey, G. (eds.): *Meaning in Mind*, Basil Blackwell, London, 1991.

Ludlow, P. & Martin, N. (eds.): *Externalism and Self-Knowledge*, CSLI Publications, Stanford, 1998.

McDowell, J.: 'De Re Senses', *The Philosophical Quarterly*, 34/136, 1984, pp. 283-294.

—————: 'Singular Thought and the Extent of Inner Space', in *Subject, Thought and Context*,
(eds.) Evans & McDowell, 1986.

McGinn, C.: 'The Structure of Content' in his *Knowledge and Reality*, Oxford University Press, Oxford, 1999, pp. 111-51.

McKinsey, M.: 'Anti-Individualism and Privileged Access', *Analysis*, 51, 1991, pp. 9-16.

—————: 'The Semantics of Belief Ascriptions', unpublished manuscript, 1998.

McLaughlin, B. & Tye, M.: 'Externalism, Twin-Earth and Self-Knowledge' in *Knowing Our Own Minds*, (eds.) Wright, C., Smith, B. & Macdonald, C., Clarendon Press, 1998, pp. 285-320.

Moore, A. W. (ed.): *Meaning and Reference*, Oxford University Press, Oxford, 1993.

Moore, J. G.: 'Misdisquotation and Substitutivity: When not to infer Belief from Assent', *Mind*, Vol. 108, 1999, pp. 335-65.

Neale, S.: *Descriptions*, MIT Press, Cambridge, 1990.

Noonan, H.: 'Rigid Designation', *Analysis*, 1979, pp. 174-182.

Peacocke, C. A. B.: 'Proper Names, Reference, and Rigid Designation' in Blackburn, S., (ed.): *Meaning, Reference and Necessity*, 1975, pp. 109-132.

Pessin, A. & Goldberg, S., (eds.): *The Twin Earth Chronicles*, M. E. Shape, New York, 1996,

Prior, A.: 'The Possibly-true and the Possible', in his *Papers in Logic and Ethics*, Duckworth, London, 1976.

Pryor, J.: *Thinking About Water*, forthcoming, 2000, pp. 1-135.

Putnam, H.: 'Is Semantics Possible?', *Metaphilosophy*, 1, no. 3, 1970.

———: 'Reference and Understanding', in his *Meaning and the Moral Sciences*, 1978, pp. 97-117.

———: *Meaning and the Moral Sciences*, Routledge & Kegan Paul Ltd., Boston, 1978.

———: *Reason, Truth, and History*, Cambridge University Press, 1981.

———: 'Is Water Necessarily H₂O?', in his *Realism with a Human Face*, Harvard University Press, 1990.

———: 'Meaning and Reference' in *Meaning and Reference*, (ed.) A. W. Moore, 1993.

———: 'The Meaning of "Meaning"' in *The Twin Earth Chronicles*, (eds.) Pessin, A. & Goldberg, S., 1996, pp. 3-52. Originally published in *Language, Mind, and Knowledge*, Minnesota Studies in the Philosophy of Science, (ed.) K. Gunderson, vol. VII, University of Minnesota Press, 1975, pp. 131-93.

Quine, W.V.O.: *Word and Object*, The MIT Press, Cambridge, Massachusetts, 1960.

———: *From a Logical Point of View*, Harvard University Press, Cambridge, Massachusetts, 1980.

Ricanati, F.: *Direct Reference*, Basil Blackwell, London, 1997.

Richard, M.: 'Direct Reference and Ascriptions of Belief', *Journal of Philosophical Logic* 12, 1983.

Rosenthal, D. (ed.): *The Nature of Mind*, Oxford University Press, Oxford, 1991.

Russell, B.: 'Knowledge by Description and Knowledge by Acquaintance' in his *The Problems of Philosophy*, Oxford, Oxford University Press, 1980.

———: 'On Denoting' in *Basic Topics in the Philosophy of Language*, (ed.) Harnish, R., 1994, pp. 161-73.

Salmon, N.: *Reference and Essence*, Princeton University Press, Princeton, 1982.

———: *Frege's Puzzle*, Cambridge, MIT Press, 1986.

———: 'How to measure the Standard Metre', *Proceedings of the Aristotelian Society*, 1988, pp. 193-217.

———: 'Illogical Belief', *Philosophical Perspectives*, 3, 1989, pp. 243-85.

Sawyer, S.: 'Privileged Access to the World', *Australasian Journal of Philosophy*, 76, 1998, 523-33.

Schiffer, S.: 'The Basis of Reference', *Erkenntnis* 13, 1978, pp. 171-206.

—————: 'Belief Ascription', *Journal of Philosophy*, no. 10, 1992, pp. 499-521.

—————: 'Descriptions, Indexicals, and Belief Reports: Some Dilemmas (But Not the Ones You Expect)', *Mind*, 104 (403), 1995, pp. 107-31.

Schwartz, 'Putnam on Artefacts' in *The Twin Earth Chronicles*, (eds.) Pessin, A. & Goldberg, S., 1996, pp. 81-8.

Searle, J.: 'Proper Names', *Mind*, Vol. 67, 1958, pp. 166-73.

Segerberg, K.: 'Two-Dimensional Modal Logic', *Journal of Philosophical Logic*, 2, 1973, pp. 77-96.

Smullyan, A.: 'Modality and Descriptions', *The Journal of Symbolic Logic*, 13, 1948, pp. 31-7.

Soames, S.: 'Review of Evans' 'Collected Papers"', *The Journal of Philosophy*, 1989, pp. 141-56.

—————: 'Substitutivity', in Thomsen (ed.) *On Being and Saying. Essays for Richard Cartwright*, Cambridge, Mass.: MIT Press, 1987a.

—————: 'Direct Reference, Propositional Attitudes, and Semantic Content', *Philosophical Topics*, 15, 1987b, pp. 47-88.

—————: 'The Modal Argument: Wide Scope and Rigidified Definite Descriptions', *Nous*, 32, 1998, pp. 1-22.

Sosa, D.: 'The Import of the Puzzle about Belief', *Philosophical Review*, Vol. 105, No. 3, 1996, pp. 373-402.

Stalnaker, R.: 'Possible Worlds', *Noûs*, 10, 1976, pp. 65-75.

—————: 'Assertion', in his *Context and Content*, 1999a, pp. 78-95.

—————: 'Semantics for Belief', 1999a, 117-29.

—————: 'On What's in the Head', in his *Content and Context*, 1999a, pp. 169-93.

—————: 'Narrow Content', in his *Context and Content*, 1999a, pp. 194-209.

- : 'Twin-Earth Revisited', in his *Content and Context*, 1999a, pp. 210-21.
- : *Context and Content*, Oxford University Press, Oxford, 1999a.
- : 'Reference and Necssity', in *A Companion to the Philosophy of Language*, (eds.) B. Hale & C. Wright, Blackwell, 1999b, pp. 534-54.
- Stanley, J.: 'Rigidity and Content' in *Logic, Language, and Reality: Essays in Honour of Michael Dummett*, (ed.) R. Heck, 1997, pp. 131-56.
- : 'Names and Rigid Designation', in *A Companion to the Philosophy of Language*, (eds.) B. Hale & C. Wright, 1999a, pp. 555-85.
- : 'Understanding, context-relativity, and the Description Theory', *Analysis*, Vol. 59, no.1, 1999b, pp. 14-8.
- Tichy, P.: Kripke on Necessity A Posteriori, *Philosophical Studies* 43, 1981, pp. 225-41.
- van Inwagen, P.: 'Indexicality and Actuality', *Philosophical Review*, 89, 1980, pp. 403-26.
- Yablo, S.: 'Textbook Kripkeanism & the Open texture of concepts', forthcoming, 2000,
- White, S.: 'Partial Character and the Language of Thought', *Pacific Philosophical Quaterly*, 63, 1982, pp. 347-65.
- Wittgenstein, L.: *Philosophical Investigations*, Basil Blackwell, Oxford, 1953.
- Woodfield, A. (ed.): *Thought and Object*, Oxford University Press, Oxford, 1982.
- Wright, C.: 'Wittgenstein's Later Philosophy of Mind: Sensation, Privacy and Intention', in (ed.) Puhl, K., *Meaning Scepticism*, de Gruyter, Berlin, 1991, pp. 126-47.
- : 'Self-Knowledge: The Wittgenstainian Legacy' in *Knowing Our Own Minds*, (eds.) C. Wright, B. Smith & C. MacDonald, Clarendon Press, 1998, pp. 13-45.
- : 'Cogency and Question-Begging: Some Reflections on McKinsey's Paradox and Putnam's Proof', *Philosophical Issues*, Vol. 10, 2000, pp. 140-63.